The Dissertation Committee for Tomas Alan Bogardus certifies that this is the approved version of the following dissertation:

AN EPISTEMOLOGICAL APPROACH TO THE MIND-BODY PROBLEM

Committee:

_____
Adam Pautz, Supervisor

_____
Michael Tye, Co-Supervisor

_____
David Barnett

_____
Daniel Bonevac

_____
Robert Koons

_____
David Sosa

AN EPISTEMOLOGICAL APPROACH TO THE MIND-BODY PROBLEM

by

Tomas Alan Bogardus, B.S.; M.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2011

For Tanja and Minna,

and our uncharted seas.

AN EPISTEMOLOGICAL APPROACH TO THE MIND-BODY PROBLEM

Tomas Alan Bogardus, Ph.D.

The University of Texas at Austin, 2011

Supervisors: Adam Pautz and Michael Tye

This dissertation makes progress on the mind-body problem by examining certain key features of epistemic defeasibility, introspection, peer disagreement, and philosophical methodology. In the standard thought experiments, dualism strikes many of us as true. And absent defeaters, we should believe what strikes us as true. In the first three chapters, I discuss a variety of proposed defeaters—undercutters, rebutters, and peer disagreement—for the seeming truth of dualism, arguing that not one is successful. In the fourth chapter, I develop and defend a novel argument from the *indefeasibility* of certain introspective beliefs for the conclusion that persons are not complex objects like brains or bodies. This argument reveals the non-mechanistic nature of introspection.

TABLE OF CONTENTS

CHAPTER ONE

Undercutting Dualism

In the standard thought experiments, dualism strikes many philosophers as true. Among these philosophers are many non-dualists. This 'striking' generates prima facie justification: in the absence of defeaters, we ought to believe that things are as they seem to be. In this paper, I examine several proposed undercutting defeaters for our dualist intuitions. I argue that these proposals all fail, either because they rest on false assumptions, require empirical evidence that they lack, or because they overgenerate defeaters. It follows, then, that our prima facie justification for dualism is as-of-yet undefeated.

1. Dualism and Non-Dualism

Naturalism is widespread in contemporary philosophy. Nevertheless, it is difficult to say what naturalism is. Perhaps it is a purely ontological commitment: all and only those entities currently posited by our best scientific theories exist, along with certain natural composites of these fundamental entities. However, given the inevitable progression of science, some naturalists are reluctant to

commit themselves to naturalism so understood. After all, in a few short decades our best scientific theories will no doubt posit different entities, in which case naturalism is currently false. Perhaps it is better to construe naturalism as a certain type of *research program*: a disposition to treat as basic all and only the evidence gained from our best scientific methods (cf. Rea 2002).

Many philosophers feel a tension between their naturalism on the one hand and their mental lives on the other hand. On the one hand, there are what we may call "naturalistically-acceptable" types of states, i.e. those state types the existence of which is well-supported by the evidence we've gained from our best scientific methods. On the other hand, there are our mental states. And these do not obviously fit into this framework of naturalistically-acceptable state types. No doubt our mental state tokens correlate very nicely with naturalistically-acceptable state tokens; the problem, if there is one, occurs at the level of *types*. With which naturalistically acceptable state types shall we identify our mental state types?

This is a difficult question with no clear answer. Nevertheless, many philosophers believe that for any mental state type M, there is some type T such that T is a naturalistically-acceptable state type (e.g. a physical brain state type,[1] or some physical-functional state type,[2]

[1] Empirical candidates for such a physical type are *displaying activity in the*

or some purely formal functional state type,[3] or some physical representational state type[4]), and M is identical with T. I will call those philosophers *non-dualists*. Other philosophers believe that there is at least one mental type that is not identical with any naturalistically-acceptable state type. I will call those philosophers *dualists*.

Unhappily, the vagueness inherent in the term "naturalism" spills over into "dualism" and "non-dualism," as I've defined them. Just what types of states count as acceptable to naturalism? I will here briefly address some common views in the literature, and say whether I take them to count as forms of dualism or non-dualism.

I take the following as paradigm cases of naturalistically-acceptable state types: physical brain state types, physical-functional state types, purely formal functional state types, physical representational state types. And so, as I understand non-dualism, it

---

pyramidal cells of layer 5 of the cortex involving reverberatory circuits (cf. Block and Stalnaker 1999), or *being cortio-thalamic oscillation* or *being C-fibers firing*.

[2] I have in mind here a chauvinistic functionalism. An empirical candidate for such a physical-functional state type would embed empirical information into our Ramsey sentence. See e.g. Block 2006.

[3] This is what Block calls the "deflationary" view: phenomenal properties are identical to some purely formal functional type, i.e. our Ramsey sentence has no empirical information.

[4] For example, a PANIC state realized in the brain that represents tissue damage as bad (Tye 1995, 2000, 2006). Or perhaps a representational state realized in the brain that represents a cluster of properties nonconceptually, which properties are suitably poised to bring about cognitive responses (Tye 2007). Presumably, properties in this cluster are not irreducibly non-physical.

clearly includes standard brain-state identity theories. Standard forms of functionalism as I understand them tend to clearly count as non-dualist theories, since they countenance in their Ramsey sentences only property types that uncontroversially count as naturalistically acceptable. And many forms of representationalism will count as non-dualist, since such theories countenance only naturalistically-acceptable state types in the contents of experiences.

Consider now a view on which mental state types are *composed* by but perhaps not *identical* with naturalistically-acceptable state types. Such a view comes in several varieties. Those varieties that say mental state types are not identical with naturalistically-acceptable mental state types count as forms of dualism, as I've defined it. Those varieties on which mental state types are identical with naturalistically-acceptable mental state types (though perhaps not those same types of which they are composed) count as forms of non-dualism.

Consider now what we might call, following Richard Spencer-Smith (1995), "radical emergentism": mental state types are novel, and no naturalistically-acceptable theory can predict or explain the appearance of them. (Here, a mental state type P is "novel" is the sense that we have P, but none of our constituents have any determinate state types of the same determinable as P.)

Again, such a view comes in several varieties. Those varieties that say mental state types are not identical with naturalistically-acceptable mental state types count as forms of dualism, as I've defined it. Those varieties on which mental state types are identical with naturalistically-acceptable mental state types (though perhaps not those same types from which they "emerge") count as forms of non-dualism.

Some may find my use of "dualism" and "non-dualism" as departing from their traditional uses in this debate. Perhaps that's right. My interest, however, is to explore and evaluate non-naturalistic theories of the mind. I intend for "dualism" to cover all and only those views which posit one of more types of properties the existence of which would be inconsistent with naturalism, even if these properties supervene with metaphysical necessity on naturalistically-acceptable properties.

2. *Dualist Intuitions*

Many dualists believe dualism because, they say, propositions that clearly entail dualism seem true to them in light of well-known thought experiments.[5]

[5] Similarly, most of us believe that knowledge is not justified true belief

For example, consider yourself with an autocerebroscope arranged in such a way that you can observe the states of your own brain and the interactions of its parts in exquisite detail, while you experience pain. Let the type-demonstrative concept THIS – deployed introspectively – refer to the horrible felt quality type of pain. Let THAT refer to whichever of the candidate naturalistically-acceptable types you care to demonstrate. Now consider each instance of the following proposition:

Possibility*:* This and that are not necessarily coextensional.

Many philosophers—including many non-dualists—have reported the seeming truth of Possibility with respect to at least some type-identity claims. For example, Saul Kripke (1972, p. 147), inspired by Descartes, says "just as it seems that the brain state could have existed without any pain, so it seems that the pain could have existed without the corresponding brain state." David Papineau (2002, 85) reports that "...it certainly *seems* possible that these properties should come apart," and admits that zombies and ghosts seem possible (ibid., 94). Christopher Hill (1996, 67) accepts the "apparent separability of pain and C-fiber stimulation." Speaking of Kripke's intuition, Colin McGinn (2003, 153) says that our correct theory "must deactivate the intuitions of contingency that surround our

---

because it sure *seems* like Gettier's Smith, who has a justified true belief, doesn't know. That's the sense of "seem" relevant to this dualist claim.

thinking about the relation between the mind and the body." And Thomas Polger (2004, 42) says "It certainly seems that my pain now could have been other than, say, activation of C-fiber #237 now… Mind-brain identity claims have the appearance of contingency."

Possibility has been widely discussed in the literature. It entails but is not entailed by a much less-discussed non-modal proposition that straightforwardly entails dualism:[6]

Non-Identity: This is not identical with that.

Many philosophers–including many non-dualists–have reported the seeming truth of Non-Identity, with respect to at least some of the type-identity claims above:

- Papineau (2002, 3): "We find it almost impossible to free ourselves from the dualist thought that conscious feelings must be something *additional* to any material goings-on... the compelling intuition that the mind is ontologically distinct from the material world... we feel it is *obvious* that conscious states are not material states."
- Daniel Dennett (1992, 27): "It does seem as if the happenings that *are* my conscious thoughts and experiences cannot be brain happenings, but must be *something else,* something caused or produced by brain happenings, no doubt, but something in addition..."

[6] We should not accept a modal-separability criterion for property distinctness. After all, there are examples of properties–ways things could be–that are distinct even though they could not fail to be coinstantiated. Triangularity and trilaterality, for example.

- Christopher Hill (2005, 153): "When I am attending introspectively to a pain, I am aware of something that appears to resist characterization in terms of neuroscientific concepts. To apply neuroscientific concepts to it would be like applying them to a patch of blue sky."
- Peter van Inwagen (2009): "[The Hard Problem is] the question: 'How could this collection of molecules actually have this kind of awareness that is my feeling of pain or orange?' And indeed I don't see how it could. In fact, it looks to me as if it couldn't, except for the fact that it does."[7]

Somewhat surprisingly, Non-Identity seems true even to hardcore eliminative materialists. According to Stephen Stich (1991, 1996), the motivating idea of eliminative materialism is that some theoretical terms fail to refer due to a *high degree of mismatch* between reality and the supposed nature of this theoretical posit. So, for example, Richard Rorty (1965) and many after him suggest that the folk theoretical expression "demon possession" fails to refer since the reality of the situation—manifestations of hallucinatory psychosis or epilepsy—is very different from the supposed nature of demon possession. Due to this high degree of mismatch (and our preference for the neuroscientific theory), Rorty and others say that "demon possession" fails to refer; really, there isn't any demon possession.

[7] Of course, here Van Inwagen just reports an intuition that this collection of molecules couldn't *have* phenomenal states. But since presumably he also thinks that this collection of molecules could token the relevant naturalistically-acceptable reductive types mentioned above, I take it that Non-Identity also seems true to him.

Now, according to Stich, the idea behind Rorty's "disappearance theory" and later eliminativist views seems to be that very many if not all folk-psychological terms (such as "pain") fail to refer, due to a high degree of mismatch between the supposed nature of these theoretical entities and reality. That is, Non-Identity seems true to these theorists, and this seeming is recalcitrant in the face of neuroscience. But given their commitment to materialism, they opt to deny that folk-psychological terms refer. For example, Paul Feyerabend (1963a, 295) says that the "usual" or "ancient" sense of the term *mental* is essentially non-materialistic, and (1963b, 54) that, on the basis of introspection, it appears that thoughts (if they exist) are very different from material processes. But he embraces reductive materialism. He therefore advocates saying there are no mental processes in this sense, and that there are no thoughts. And so it is in large part the seeming truth of Non-Identity– i.e. his strong dualist intuition–that drives him to this conclusion.

So, Possibility and Non-Identity seem true to wide variety of philosophers, including many non-dualists.[8] Of course, in that respect these propositions are not unique: there are very many propositions concerning pain that seem obviously true in this way.

[8] Eliminativists count as non-dualists on my definition, since they claim that there are no mental state types, and so they claim it's false that there is some mental state type that fails to be identical with some reductive type. That is, they claim dualism is false.

For example, it seems obvious that pain is not pleasure, and that a state of affairs could include pain and yet not include anything circular. Yet such propositions do not straightforwardly entail dualism. By contrast, Non-Identity clearly entails dualism,[9] and Possibility clearly entails Non-Identity.[10] And, to repeat, these propositions seem true even to many non-dualists. They also seem true to me. Perhaps they seem true to you as well.

This dissertation examines a novel argument for the conclusion that we should be dualists. Since the central argument crucially depends on the standard methodology of philosophy, I will call it "the Methodological Argument."

David Chalmers (1996) famously argues for dualism with respect to phenomenal mental state types – for example, the awful felt quality type of pain.[11] Chalmers' argument for dualism is widely regarded as the best contemporary argument for dualism. However,

---

[9] More precisely, the proposition that *this* is not *that,* or *that,* or *that…* (where each deployment of THAT refers to a distinct member of the domain of candidate reductive types and every member of that domain is referred to by an instance of THAT) both intellectually seems true and clearly entails dualism.

[10] Non-Identity by itself entails that non-dualism–as I have defined it–is false, but is consistent with a Nagelian-style primitivism, according to which mental phenomena supervene on physical phenomena in virtue of some metaphysically necessary relation which is not identity. As I have defined dualism, Nagelian primitivism is a form of dualism. Possibility is stronger, entailing that there is a metaphysically possible world in which *this* and *that* come apart, and therefore it entails that Nagelian primitivism is false.

[11] Pinch yourself hard. *That* type of feeling.

his argument relies on two-dimensional semantics, peculiar notions of conceivability, and the claim that the conceivability of zombie worlds entails their possibility.[12] Critics have attacked all three of these parts of the argument.

To its credit, the Methodological Argument that I examine in this dissertation sidesteps these controversies that plague Chalmers' argument. Instead of relying on two-dimensional semantics, the Methodological Argument uses the widely accepted standard justificatory procedure of philosophy. And instead of using Chalmers' notions of conceivability, the Methodological Argument uses a minimal and uncontroversial notion of *intellectual seeming* (more on this below). Finally, the Methodological Argument does not require that conceivability entail possibility. For these reasons, the Methodological Argument that I consider in this dissertation is far more threatening to non-dualism than Chalmers' argument.

3. Intellectual Seemings

The verb "to seem" has many senses. In the doxastic sense, it *seems* like the Republicans will pick up some congressional seats in 2012. In the visual sense, it *seems* like the stick in that glass of water is

[12] Though Chalmers (2010) provides an intentionally non-technical argument for dualism that relies only on considerations about structure and function.

bent (though I don't believe that it is). In a third sense, the naïve comprehension axiom *seems* true, even though I don't currently believe it since I've seen compelling proofs of its falsity. It is this third sense of "to seem" that the Methodological Argument uses. I will call this sense the "intellectual" sense, and I will speak of "intellectual seemings." This is a type of conscious episode, distinct from occurrently believing and sensorily experiencing, in which a proposition strikes one as obviously true, i.e. in which one apparently just sees the truth of a proposition.[13]

Like much else in philosophy (e.g. causation, numbers, induction, universals, consciousness), intellectual seemings may appear mysterious or exotic upon reflection. But at the same time, they're as common as dirt. For example, if you learn that Alan is taller than Bob, and that Bob is taller than Charles, the proposition *Alan is taller than Charles* will – upon consideration – seem true to you in this sense of "seem." And if you learn that I'm angry with a man named "Smith," it will seem to you – upon consideration and again in this intellectual sense of the verb "seem" – that the man Smith is not an even integer. And very many of our cherished logical rules seem true to us upon consideration in this same sense of "seem," from modus ponens to De Morgan's laws.

[13] See Conee 1998 for more on "seeing the truth" of a proposition.

Intellectual seemings are distinct from both of Chalmers' (2010, chapter 6) varieties of conceivability featured in his most well-known argument for dualism. There, he distinguishes negative conceivability – where *p* is negatively conceivable for S just in case *p* cannot be ruled out a priori by S – from positive conceivability – where *p* is positively conceivable for S just in case S "can (in principle) clearly and distinctly imagine a situation" in which *p* is true. The Methodological Argument departs from Chalmers' argument in this important respect: for *p* to intellectually seem true is not for *p* to be conceivable, either positively or negatively. Let me explain.

A proposition can intellectually seem true without it being negatively conceivable. For example, the naïve comprehension axiom intellectually seems true to me, and yet I *can* rule it out a priori with the aid of well-known proofs. Also a proposition can be negatively conceivable, and yet not intellectually seem true. For example, I cannot rule out Goldbach's conjecture a priori, and yet it does not intellectually seem true to me. Therefore intellectual seeming is not negative conceivability. Neither is it positive conceivability, since it intellectually seems to me that the number 17 is prime, even though I cannot even in principle form a mental image of the number 17 at all,

and therefore I cannot form a mental image in which the number 17 is prime.[14]

To recap, intellectual seemings are common and familiar to us. So their existence should be uncontroversial. And a proposition may intellectually seem true to S without S thereby believing it or even being disposed to believe it, as happens with the naïve comprehension axiom.[15] And, finally, intellectual seemings are distinct from both of Chalmers' varieties of conceivability.

4. The Methodological Argument for Dualism

The standard justificatory procedure in philosophy (cf. Bealer 1992 and 1996) counts all the following as prima facie evidence:

---

[14] **Objection**: One can very easily form a mental image of the *numeral* "17". And that's in fact how one positively conceives of propositions involving the number 17. **Response**: In that case, positive conceivability is still not intellectual seeming, since *17 is red* will then be positively conceivable (just imagine a red-colored "17") and yet will not intellectually seem true. **Another Objection**: One may very easily form a mental image of a situation in which *17 is prime* is true. That proposition is true in every possible situation. So imagine any possible situation you like, and you've succeeded. **Response:** In that case, positive conceivability is still not intellectual seeming, since then either Goldbach's conjecture or its negation will be positively conceivable, and yet neither intellectually seems true.

[15] On a liberal sense of "disposed to believe," I *am* disposed to believe the naïve comprehension axiom, since I would believe it given the right evidence base (namely, given its seeming truth and a lack of defeaters that I actually have). But even on this liberal sense of "disposed to believe," dispositions to believe are not intellectual seemings: I am disposed to believe (in this liberal sense) that John McCain is President, since I would believe it given a certain evidence base. And yet that proposition doesn't intellectually seem true to me.

experiences, observations, testimony, and – importantly— intellectual seemings such as those elicited by dualist thought experiments. Examples of this procedure abound: Gettier's refutation of K=JTB, Chisholm's perceptual relativity refutation of phenomenalism, Putnam's Spartan-pretender refutation of behaviorism, all the various twin-earth examples, Burge's arthritis example, multiple realizability, etc. These examples all involve the evidential use of intellectual seemings, which some philosophers call "intuitions."

So widespread philosophical practice and methodology supports the view that intellectual seemings confer at least prima facie, defeasible justification. I will take it that this widespread practice is correct: intellectual seemings give us defeasible justification.[16] How are intellectual seemings defeasible? Pollock (1974) distinguishes between two types of defeaters: rebutting and undercutting. Rebutting defeaters, in this case, are arguments or evidence for the negation of Possibility or the negation of Non-Identity.

An undercutting defeater, on the other hand, is more difficult to characterize. Pollock's classic analysis would have it that an

[16] Note that this claim is weaker than Michael Huemer's (2007, 30) phenomenal conservatism, which entails that every kind of seeming confers prima facie justification.

undercutting defeater, in this case, is a reason to think that the following subjunctive conditional is false: dualism would not seem true unless it were true. This analysis is not uncontroversial, however. Whatever the right analysis of undercutting defeat turns out to be, an undercutting defeater attacks the relation between one's belief and the grounds on which one holds the belief. In this dissertation, we will rely only on the following necessary condition for an undercutting defeater: some consideration is a successful undercutting defeater for a subject S's rational belief B only if upon learning of the consideration S can no longer rationally persist in holding B. We will test proposed undercutting defeaters by seeing whether such a consideration entails that a subject could not rationally persist in her belief.

As it stands, those of us to whom Possibility and Non-Identity intellectually seem true are prima facie justified in believing them. Ultima facie justification results from searching for defeaters to an extent that satisfies our relevant epistemic obligations and finding none. Furthermore, not only are we permitted to believe Possibility and Non-Identity in the absence of defeaters, but we epistemically *ought* to. For it is plausible that for any subject S, if S is permitted to believe that *p* and (i) it's not the case that S is permitted to suspend belief that *p*, and (ii) it's not the case that S is permitted to believe that *not-p*, then S *ought* to believe that *p*. It's always nice, after all, to gain

one more true belief. And if you find yourself in a situation in which you are epistemically permitted only to believe that *p*, then you ought to believe that *p* (that's your best bet, so to speak). Therefore we epistemically ought to believe Possibility and Non-Identity unless we have or gain access to defeaters for them.

Progress on the mind-body problem might be made, therefore, by establishing whether there are any defeaters for the seeming truth of Possibility and Non-Identity. Let's call this "the Methodological Argument for Dualism." In the present chapter, our project is to examine several proposed undercutting defeaters. We will ask whether the type of consideration proposed entails that a subject could not rationally persist in her belief. We will find that, in each case, the answer is 'no'.

5. Undercutting Dualism

5.1. Unreliability with respect to seemings generally?

If I examine the track record and learn that the majority of beliefs produced by my memory have been false, I thereby gain a defeater for any belief I currently hold on the basis of memory. Similar considerations apply to the seemings relevant to the Methodological Argument. One might suspect that if I examine the

track record and learn that the majority of beliefs I have held on the basis of seemings have been false, I thereby gain a defeater for any belief I currently hold on the basis of a seeming, e.g. my belief in Possibility and my belief in Non-Identity.

I have met people with a high degree of confidence that an examination of the track record of my beliefs based on seemings would in fact reveal that seemings are unreliable. These people have claimed that tables seem not to be mostly empty space, that the Earth seems to be flat, and that it seems possible for a world to have water but no H_2O. Yet, these people continue, science has taught us that these seemings are mistaken. And, they add, there are many other cases in which things have been other than how they seemed to be. So, they conclude, we have a defeater for any beliefs that we currently hold on the basis of a seeming, including belief in Possibility and belief in Non-Identity.

5.2 Response

Possibility and Non-Identity seem true in the intellectual sense of "seem." This sense, recall, is distinct from the doxastic sense, i.e. inclinations to believe. Consider, for example, the naïve comprehension axiom, which continues to seem true to me in the sense relevant to the Methodological Argument, and yet which I am

not inclined to believe (in a non-trivial sense of "inclined to believe"). Intellectual seemings, recall, are also distinct from visual seemings, i.e. that sense of "seem" in which it seems to me that there is a white surface with black markings before me.

So, my beliefs in Possibility and in Non-Identity are based on intellectual seemings. Yet, contrary to the above objection, it would not automatically follow from the fact that seemings in general are unreliable that *intellectual* seemings are unreliable. Similarly, one's vision might be impeccable even if every other sense modality were completely unreliable. Though, on the whole, such a person's beliefs based on sense perception might be unreliable, neverthless her beliefs based on vision may be completely trustworthy. The same may go with seemings in general and intellectual seemings in particular. So the objection must show that our intellectual seemings do not track the truth.

But examples of mistaken seemings – in the sense of "seeming" relevant to the Methodological Argument – are not nearly so numerous as some may believe. I contend that the seemings involved in the above claims – namely the claims that tables seem not to be mostly empty space, that the Earth seems flat, and that it seems possibly for a world to have water but no H_2O – are either not intellectual seemings and so not by themselves damaging to the

Methodological Argument, or are intellectual seemings but are not mistaken. Let me first discuss the case of tables and the shape of the Earth.

It's true that tables are mostly empty space, and it's true that most people unenlightened by science fail to believe that they are, and in fact may believe that they are not. I was once a member of that group. But, in my case at least, the belief that tables are not mostly empty space was not based on an intellectual seeming. Nor, I think, did it even *visually* seem to me that tables are not mostly empty space. As we know, tables that are mostly empty space can look the same as tables that are not mostly empty space. And so our visual seemings that attend the viewing of tables are consistent with both hypotheses; our visual experience while viewing tables is therefore not aptly described as *seeming not to be mostly empty space*. No, it is rather that tables, upon viewing, fail to seem to be mostly empty space. The inclination to believe that tables are not mostly empty space is not based solely on the visual seeming, but rather at least partly on other considerations, perhaps, for example, simplicity.

Put another way, the commonsense notion of solidity is something like *having no visible holes,* or *impenetrable or resistant to transformation within some range of forces*. A more sophisticated notion of solidity is that of having no holes whatsoever, visible or otherwise.

I take it that what we mean when we report that a table seems solid, on the basis of tactile or visual inspection, is that it seems solid in the ordinary sense of solidity. That the table is solid in the more sophisticated sense simply goes well beyond our visual or tactile evidence, since, as we know, a cloud of microparticles would feel and look just the way that a uniformly dense table would feel and look. Though, perhaps because the hypothesis is simpler and hence more attractive to minds predisposed toward simplicity, we may have once been inclined to believe that ordinary physical objects are uniformly dense on the basis of tactile and visual inspection, we unhesitatingly gave that up when science uncovered evidence that ordinary physical objects are actually composed of clouds of microparticles. Science delivered undercutting defeaters for our inclination to believe that ordinary physical objects are uniformly dense, on the basis of tactile or visual inspection. But this in no way impugns the integrity of our *intellectual* seemings. They simply weren't involved.

Similar considerations apply, I think, to the case of the shape of the Earth. There may have been a point when, as a child, I looked at the horizon and falsely believed that the Earth was flat. But if I did, my belief was not based on an intellectual seeming – it's not as though the proposition that *the Earth is flat* seemed true in the way that *no prime minister is a prime number* seems true. Nor, I think, was it

even based solely on the visual seeming. As we know, a flat Earth and an extremely large spherical Earth may look the same to one who stares off into the horizon. Whatever my visual experience of the shape of the Earth was while staring at the horizon as a child, it is not best described as *seeming to be flat* or *seeming not to be spherical*. No, it is rather that the Earth, upon viewing, failed to seem spherical.[17] I was merely inclined to believe, perhaps for reasons of simplicity, that it was flat. In any event, these are not cases in which I formed false beliefs on the basis of intellectual seemings (nor even solely on visual seemings). And so neither these examples nor the numerous examples like them can furnish me with a defeater for my belief in Possibility and my belief in Non-Identity, beliefs which are both based solely on intellectual seemings.

Things are a bit trickier when it comes to water. Here, I think, one's belief that it's possible for a world to contain water but no H_2O *is* based on an intellectual seeming. But this intellectual seeming is not misleading, despite the fact that water is necessarily identical with H_2O. I will turn to this issue in the next section.

[17] I take it that Ludwig Wittgenstein made a similar point, in a discussion with G.E.M. Anscombe. Wittgenstein wondered aloud: "It's always puzzled me why people believed that the Sun went around the Earth." Anscombe replied: "Isn't it obvious? It's because it *looks* like the Sun goes around the Earth." "And how would it look if the Earth went around the Sun?" Wittgenstein replied.

5.3 Unreliability with respect to a certain kind of intellectual seeming?

The previous section argued that we do not have a track record of widespread error when it comes to our beliefs based on intellectual seemings. So that consideration cannot furnish us with an undercutting defeater for our dualist intuitions. But suppose I examine the track record and learn that I am unreliable with respect to a certain *kind* of intellectual seeming. Suppose, for example, that when it comes to the mathematics of infinity, very many propositions that intellectually seem true to me are actually false. (For example, it occasionally strikes me as true that there are more natural numbers than even numbers. But this is false. And it occasionally strikes me that 1 is greater than the infinite decimal 0.999…. This too is false.) I may thereby gain a defeater for any belief I currently hold about the mathematics of infinity on the basis of an intellectual seeming. Similarly, suppose I examine the track record and learn that, when it comes to non-tautologous identity statements involving natural kind terms, very many propositions that intellectually seem true to me are actually false. I may thereby gain a defeater for any belief I currently hold about such identity statements on the basis of intellectual seemings. And this would include Possibility and Non-Identity.

Some philosophers believe that we are in this exact situation. For example, they say, water = H_2O. And yet, they say, there is an

appearance of contingency here. It seems possible for water and H_2O to come apart, and (so) they just seem different. And this holds generally with respect to non-tautologous identity statements involving natural kind terms. Therefore, they conclude, considerations like these furnish us with an undercutting defeater for many beliefs we hold regarding those kinds of identity statements. Among these beliefs that are undercut, they insist, are Possibility and Non-Identity.

5.4 Response

What some *report* as the seeming possibility of a world with water but no H_2O is, indeed, an intellectual seeming. But this intellectual seeming is not misleading; it is merely misreported. Let me explain.

The proposition that

(1) Water is H_2O

says of one natural kind–known as both "water" and "H_2O"–that it is self-identical. If this is so, then since (1) is true it is necessarily true. The objector urges that in some sense (1) seems possibly false. That is, the objector urges that the following proposition seems possibly true:

(2) Water is not H_2O

But since (1) is necessarily true, (2) is necessarily false. And so the objector may conclude that when it comes to the modal status of identity claims involving natural kind terms, intellectual seemings are unreliable guides to truth. Since my belief in Possibility and Non-Identity are based on intellectual seemings, the objector concludes that I have an undercutting defeater.

But I and many others think the objector suffers from proposition-confusion.[18] She mistakenly takes the sentence "Water is H_2O"to express something like one of these propositions:

(3) The watery stuff of our acquaintance is H_2O

(4) "Water" and "H_2O" corefer

And *these* propositions seem possibly false. But (3) and (4) are indeed as they seem: they are (metaphysically) possibly false.

A person who thought that the sentence or utterance "Water is H_eO" expresses (3) or (4) would likely issue the report that "it seems possible that water isn't H_2O," since of course (3) and (4) are contingent propositions. Perhaps such a person thinks that natural

[18] This is, I take it, the common interpretation of Kripke. Others have followed. For example, Michael Tye (1986, 5) says "If…a man without scientific knowledge claims to be imagining that gold has atomic number 80 (rather than its actual 79) what I think we would say he *really* imagines is that some substance with the superficial observable qualities of gold has atomic number 80 (rather than 79), and that is something quite different."

kind terms merely abbreviate non-rigid definite descriptions, and so she thinks that the sentence "Water is H_2O" expresses (3). What seems true to her isn't that (1) is possibly false, but that (3) is possibly false.

Or perhaps the person is confusing word and object, and what seems true to her isn't that (1) is possibly false, but that (4) is possibly false. But in each case, then, intellectual seemings have not led this person astray. She has not erred with respect to the modal statuses of propositions. Rather, she is misunderstanding the semantics of the sentence, thinking that a *sentence* (namely "Water is H_2O") expresses a contingent proposition [namely (3) or (4)] when really it expresses a necessary proposition [namely (1)]. In each case, propositions have the modal status that they seem to have, and so we have no defeater for intellectual seemings generally.

To put it another way, Kripke did not teach us that certain propositions that appeared contingent were really necessary. He did not point out a *modal* illusion that we suffer from. Rather, Kripke taught us that proper names and natural kind terms function as rigid designators, not as disguised non-rigid definite descriptions. He drew our attention to certain sentences or utterances that we had thought expressed contingent propositions. Upon closer examination we came to see that proper names and natural kind terms are rigid

designators, and hence that these sentences or utterances express necessary propositions. Thus we are subject to a *semantic* illusion, not a *modal* illusion. But then this is not a case in which intellectual seemings have led us astray with respect to identity statements, and therefore this cannot be marshaled in support of non-dualists who claim that in the case of dualism we are suffering from a modal illusion.

So far, this is the standard Kripkean story about proposition confusion. Not all philosophers have found this convincing. For example, Scott Soames (2006) argues against what he calls the *Coherent Conceivability Thesis*,[19] which goes roughly like so:

> Apart from confusion about what we are conceiving, coherent conceivability is a reliable guide to genuine (metaphysical) possibility. If we can coherently conceive – without confusion of the sort discussed earlier in this section -- of a world in which p is true, then there are genuine (metaphysically) possible worlds in which p is true.

If Soames is right, then the dualist may still be on the hook. For, as Soames would have it, it may still be that we shouldn't trust our intuitions when it comes to identity claims involving rigid

[19] As I say, the Methodological Argument I am considering eschews talk of conceivability in favor of intellectual seemings. Soames, however, targets conceivability. In what follows, I will reconstrue Soames' criticisms so that they apply to the Methodological Argument.

designators, *even if* we take care to avoid confusing rigid designators with their associated reference-fixing descriptions. There may be other confusions that one could make, according to Soames, confusions we are in fact apt to make. This would spell trouble for our dualist intuitions.

Soames' argument against the Coherent Conceivability Thesis goes as follows. He shows that it is inconsistent with a widely-accepted Kripkean thesis about the essentiality of origins for material objects. Soames asks us to consider this sentence:

> (Sperm&Egg) "I came from a sperm and an egg (if I exist at all)."

In uttering this sentence, Soames says, I identify the referent of 'I' directly, without detour through nonrigid descriptions. Hence, no confusion of rigid designator with nonrigid reference-fixing descriptions threatens. If the Coherent Conceivability Thesis were correct, this should make for a close connection between apparent possibility and genuine possibility. But, according to Soames, it doesn't. Though the proposition I use the sentence (Sperm&Egg) to entertain and assert is (assuming Kripke's own essentiality of origin thesis, according to Soames) necessary, a world in which I do not come from a sperm and egg is apparently possible. Thus, we have a counterexample to the Coherent Conceivability Thesis: an example of

a proposition that is apparently possible yet genuinely impossible, and this failure in our modal intuitions cannot be chalked up to a confusion between rigid designators and nonrigid reference fixing descriptions.

Of course, this objection assumes that I am identical with (or otherwise necessarily connected with) my body, i.e. that material object with origins in a sperm and egg. Given the necessity of identity and Kripke's essentiality of origins thesis, it would follow that I could not have existed but with origins in a sperm and an egg. Naturally, many dualists will not accept this assumption. In fact, they may take the apparent possibility of a world in which one does not have his or her origins in a sperm and an egg as a compelling argument against the view that one is identical with (or otherwise necessarily connected with) one's body or brain.

But Soames' argument can be recast using a different kind of rigid designator, a demonstrative like "that." Consider this sentence:

> (Table) "That (pointing to the wooden table before me) came from wood."

A revised argument might go as follows: There is an apparently possible world in which the proposition expressed by (Table) is false. Yet it is necessarily true since it is actually true and Kripke's essentially of origins thesis is correct. And in this case the

demonstrative "that" refers directly, not via a nonrigid reference-fixing description. Therefore, the Coherent Conceivability Thesis is false.

I do not find this argument compelling. Three replies suggest themselves. First, I confess that a world in which *that* (i.e. the wooden table) did not come from wood does not strike me as apparently possible. I take it I'm not alone in this, and this is why Kripke chose a wooden table as his example when motivating the essentiality of origins thesis. So the first premise of the revised argument is false. Soames' premise might be more plausible if one were ignorant of the table's composition. Even then, however, it would not strike me as apparently possible that *this* (pointing to the table before me) might not have come from wood. I would be agnostic on the question.

Secondly, if one accepts that material objects can survive a complete replacement of their parts, then there are good grounds to suppose that Kripke's essentiality of origins thesis is false. David Barnett (2005 and ms.) gives counterexamples to both the claim that material objects have their origins essentially in the hunk of matter they actually came from, and also the claim that material objects have their origins essentially in the type of substance that they actually came from. The second kind of counterexample suffices to refute both claims. One such counterexample goes roughly like this. Call

this table before me *Tab*. Suppose Tab has *n* number of wooden parts. Suppose further that, prior to having been assembled into a table, Tab's *n* building blocks had undergone gradual and complete replacements of their original wood with plastic. The same carpenter who built Tab had then assembled the very same *n* building blocks, according to the very same plan, at the very same time, into a table. Intuitively, Barnett says, the carpenter would have built Tab. (Of course, only people who believe that material objects can survive complete part replacement will find this intuitively attractive. I am not among those people.) If Barnett is right, then Kripke's essentiality of origins thesis is false and so the revised argument inspired by Soames will fail.

Finally, consider the claim that in the sentence (Table) the demonstrative "that" refers directly. Whatever it means for "that" to refer directly, it must be consistent with the fact that demonstratives have modes of presentation attached with each deployment. That is, whenever we deploy a demonstrative, there is in fact some nonrigid description that we associate with the demonstrative. Perhaps the demonstrative does not *refer* via the description, whatever that might mean. But there is no doubt that the description exists and could easily be called to the mind of whomever deploys the demonstrative. In the case of (Table), the nonrigid description we associate with "that" is something like "the table before me, having such and such

manifest properties." But such a description is poised to be confused with the rigid designator "that" in the same way that we are apt to confuse "Hesperus" for "the evening star." But then we will not here have a counterexample to the Coherent Conceivability Thesis, since we will not have a case of an impossibility the coherent conceivability of which cannot be explained as an instance of proposition confusion. And thus the revised argument against our dualist intuitions fails. In conclusion, then, the prima facie justification of Possibility and Non-Identity remains undefeated.

5.5. Fallacious operator shifts, perhaps?

René Descartes claimed to see, "clearly and distinctly," that his essence did not include spatial extension, and so that he—in contrast to his body—could exist while no material objects existed. Antoine Arnauld suggested that perhaps Descartes was suffering from a failure of imagination or rational insight. Perhaps, Arnauld thought, it wasn't that Descartes clearly saw that his essence did not include spatial extension, but rather that he failed to see that it did. And so, perhaps it was not that Descartes clearly saw the possibility of his existing while no material objects exist, but rather that he merely failed to see the impossibility. By way of example, Arnauld pointed out (in Descartes 1984, 141-2) that one may see for certain

that the angle in a semi-circle is a right angle, and yet may doubt, or not be certain, or not understand that the square on the hypotenuse is equal to the squares on the other two sides. And yet it does not follow that having the square of the hypotenuse equal to the squares of the other two sides is not essential to this right triangle. I take the suggestion from Arnauld to be that Descartes may have been the victim of a fallacious operator shift, moving hastily from the truth that *p does not seem impossible* to the conclusion that *p seems possible*.

Michael Tye (1986) offers a similar suggestion:

> Where I suggest we go wrong in our thought experiments is in the belief that if it seems to us that we have imagined things *A, B, C, . . .* occurring together in some possible world W_n it automatically follows that we have really done so. …[W]e may have succeeded in imagining all of *A, B, C,...* but not together in a single possible world

Tye's worry, I take it, is that one may mistake the possibility of *A* and the possibility of *B* for the possibility of *A*&*B*, thereby committing an elementary modal fallacy.

Similar concerns can be urged against Possibility and Non-Identity. For Possibility, an Arnauldian proposal would be that perhaps we mistake our failure to see the impossibility of *this* without *that* for successfully seeing the possibility of that state of affairs. And Tye's suggestion would be that perhaps we mistake our

seeing that a possible world includes *this* and another possible world doesn't include *that* for seeing that a single possible world includes *this* but not *that.*

As for Non-Identity, a non-dualist who takes this line has two options. She could say that, after fallaciously inferring Possibility, the dualist goes on to validly infer Non-Identity. Or, the non-dualist could say that perhaps Non-Identity is itself the product of a fallacious operator shift. David Armstrong (1968, 48-49) puts the proposal this way: "It can… be suggested by the Materialist that we tend to pass from something that is true: *I am not introspectively aware that mental images are brain processes* to something that is false: *I am introspectively aware that mental images are not brain processes.*"

And so perhaps, as these authors suggest, I am the victim of some fallacious operator shift, hastily passing from the proposition that *I can't see that p is impossible* to the proposition that *I can see that p is possible,* or from the proposition that *I can see that p is possible and that q is possible* to the proposition that *I can see that (p&q) is possible*, or from the proposition that *I am not aware that p* to the proposition that *I am aware that not-p*. The question before us is, do any of these proposals furnish us with an undercutting defeater for our dualist intuitions?

5.6. Response

No, these suggestions do not provide any undercutting defeaters for Possibility and Non-Identity, for two reasons. First, merely pointing out the *possibility* of a fallacious operator shift – i.e. merely *suggesting* that this occurs – is not in general sufficient to undercut beliefs held on the basis of intellectual seemings. That type of consideration does not in general entail that a subject cannot rationally persist in her belief. And so these proposals fail to meet the necessary condition on undercutting defeat mentioned above.

Consider the seeming truth of the proposition that the Prime Minister is not a prime number. What could be more obvious than that? But now consider this suggestion from a niggling skeptic: "You mistake failing to see that the Prime Minister is a prime number for successfully seeing that he isn't." Surely this bare possibility–this unsubstantiated suggestion–does not defeat your belief that the Prime Minister is not a prime number.

Secondly, if the *mere possiblity* of a fallacious operator shift were sufficient to defeat intellectual seemings, then all of our beliefs based on intellectual seemings would be undercut, including Arnauld's, Tye's, and Armstrong's beliefs that the operator shifts they mention are fallacious. After all, for example, Tye *may* unconsciously mistake his failure to see the validity of

$((\Diamond p \& \Diamond q) \rightarrow \Diamond(p \& q))$ for successfully seeing its invalidity. But if those beliefs are undercut, i.e. if neither they nor we can justifiably believe that these operator shifts are as fallacious as they seem, then their proposals lose all force.

And so it seems perfectly rational to persist in believing Possibility and Non-Identity on the basis of their seeming truth, even upon gaining the belief that this seeming *may* be the result of an unconscious fallacious operator shift. Things would be much different if we had solid empirical data that we actually *are* victims of a subconscious fallacious operator shift. But, as the proposals stand, they do not provide us with an undercutting defeater for the seeming truth of Possibility and Non-Identity. So let us now turn to the next proposal for an undercutting defeater.

5.7. Dual-Process Cognition

In a forthcoming paper, Fiala et al. propose that our dualist intuitions arise from dual-process cognition. In contrast to the previously discussed suggestions, this proposal actually has substantial empirical evidence in its favor. The idea is that humans have a "low-road" cognitive process for attributing mental states. This process is quick, automatic, unconscious, associative, heuristic-based, computationally simple, evolutionarily old, domain-specific

and non-inferential. It is triggered by simple, surface level features, e.g. having eyes, appearing to behave in a contingently interactive manner, and displaying distinctive (non-inertial) motion trajectories. In addition to this "low-road" process, humans come equipped with a "high-road" cognitive process for attributing mental states. This process is relatively slow, controlled, introspectively accessible, rule-based, analytic, computationally demanding, inferential, domain-general, and voluntary.

Usually, these two processes issue harmonious verdicts. However, this is not so in dualist thought experiments, according to Fiala et al. In dualist thought experiments, we consider the mass of gray matter that composes the human brain. If we are non-dualists, we might deduce from our internalized theoretical beliefs that the right kind of naturalistically-acceptable state type is also a certain type of conscious experience. Or, if we are dualists, we presumably use this high-road cognitive process merely to entertain the hypothesis that a certain kind of naturalistically-acceptable state type is a certain kind of conscious experience.

At the same time, our low-road cognitive process does *not* issue the verdict that a certain kind of brain state is a conscious experience, since we do not categorize the eye-less, behavior-less, motionless brain as an agent. This low-road cognitive process fails to

make any attributions of consciousness; it is just silent on the question. We fail to have a quick, automatic 'gut-feeling' that the reductive state type in question is a certain type of conscious experience. But it would be hasty to conclude on this basis that Possibility is true: from a failure for type X to seem identical with type Y, it hardly follows that X and Y are not necessarily coextensional. Neither should we conclude on this basis that Non-Identity is true: from the fact that type X fails to seem like type Y it doesn't follow that X is not identical with Y.

Let's grant that this is all correct: we do have these two processes for attributing mental states, and only one is active during standard dualist thought experiments. Would it follow that we have an undercutting defeater for our dualist intuitions?

5.8. Response

No, it would not follow. The proposal overgenerates defeaters. If this suggestion from Fiala et al. required that we be suspicious of our dualist intuitions, it would also require that we be suspicious of perfectly mundane intuitions that we know are above reproach. For example, I look down and notice that a couple of my floorboards are misaligned. I consider the proposition that *being a misaligned floorboard* is identical with the feeling of pain. I

immediately dismiss the suggestion out of hand. *Of course* the property of being a misalingned floorboard is not identical with pain: the two properties are not coextensive, and they are just obviously non-identical. These 'dualist' intuitions, I take it, are completely above board. We should not be suspicious of them.

However, whatever disharmony occurs between my dual cognitive processes in the standard dualist thought experiments also occurs in the case of the misaligned floorboards. My high-road cognitive processes entertains a certain identity claim: being a misaligned floorboard just is pain. My low-road cognitive process does not consider the eye-less, behavior-less, motionless floorboards to be agents. That low-road process therefore does not attribute consciousness to the floorboards. If this type of disharmony between my dual cognitive processes is sufficient to undercut my dualist intuitions, it should also undercut my floorboard intuitions. But since I can clearly rationally persist in my belief that misaligned floorboards are not conscious, this proposal from Fiala et al. does not undercut my dualist intuitions.

This suggests that there is more to the story that Fiala et al. recognize. Perhaps I have some other cognitive process that issues verdicts about what types of states *could not be* conscious states. For all Fiala et al. tell us, this third process might be perfectly reliable,

and it may be operative in the case of the floorboards and also in the standard dualist thought experiments.

5.9. Duped by our concepts?

Let's now discuss what I believe to be the most popular and plausible attempt to undercut dualist intuitions. Many philosophers think that reflection on the concepts that figure in Possibility and Non-Identity will furnish us with an undercutting defeater for each.[20] These philosophers typically claim that the culprit is a lack of a priori entailment relations between some relevant concepts. Typically, they say that there are no substantive a priori ties between our phenomenal and physical (or physiological) concepts. That is, there are no non-tautalogous a priori knowable inferences from thoughts containing phenomenal concepts to thoughts containing physical (or physiological) concepts, and vice versa. You will find this thought in Block (2006, 53), Hill (1996, 75), Loar (2003, 115-6), Nagel (1998, §4), Papineau (2002, 87), Tye (1999, 715), and many others.

[20] In what follows, the philosophers I discuss typically do not construe dualist intuitions as crucially involving only demonstrative concepts like THIS and THAT. However, I believe that they would mean what they say about specifically phenomenal and physical concepts – like PAIN and C-FIBERS FIRING – to apply, for even stronger reasons, to stripped down demonstrative concepts.

So the suggestion is that the concepts that figure crucially into Possibility and Non-Identity are, from the armchair at least, really just silent with respect to each other. It is not Possibility itself that intellectually seems true. Rather, we really only *fail* to see the impossibility of *this* without *that*. And it is not Non-Identity itself that intellectually seems true. Rather, we really only *fail* to see that *this* is identical with *that*. And of course, as Arnauld and Armstrong pointed out, these are bad grounds on which to believe Possibility and Non-Identity.

Now, in order to improve on the mere *suggestions* from Arnauld and Armstrong, and in order actually get an undercutting defeater, it must be the case that, in the dualist thought experiment, we are in fact (and not merely possibly) disposed to believe dualism on these poor grounds. This, I take it, is why Nagel (1998, §4) says that "our concepts fail to reveal a necessary connection, and we are tempted to conclude to the absence of any such connection." Elaborating on Nagel's proposal, Hill (1996, 75-8) asserts that we are *in fact* endowed with an unreliable psychological mechanism the function of which is to churn out the belief that the referents of any two concepts C1 and C2 could come apart, whenever there are merely no substantive a priori ties between C1 and C2 (and no immediately accessible sufficient a posteriori reasons to think that C1 and C2 corefer).

I interpret Nagel and Hill as asserting something like the following:

> (5) For any concepts C1 and C2 and normal human subject S, if C1 and C2 have no substantive a priori ties, and if S lacks sufficient a posteriori reason to think C1 and C2 corefer, then it won't intellectually seem metaphysically impossible to S for the referents of C1 and C2 to come apart, and it will seem to S on this basis that it's metaphysically possible for the referents of C1 and C2 to come apart.

If (5) is true, and if its antecedent is met, then the seeming truth of dualism is not a reliable indicator of or does not warrant belief in its actual truth. It would be false that dualism wouldn't seem true unless it were true. This proposal, therefore, would give us an undercutting defeater for Possibility (and Non-Identity, which, presumably on this view, we infer from Possibility). However, if (5) is false, then this proposal fails to deliver an undercutting defeater.

5.10. Response

The proposal rests on the assumption that (5) is true. However, (5) is false, and so this proposal fails to deliver an undercutting defeater. The second "seem" in the consequent of (5) may refer to an intellectual seeming, or it may refer to a disposition

to believe. Here's a counterexample to (5) on both interpretations: Something somewhere is named "Chomolungma," but I am not telling you what or where it is. You now have the concept CHOMOLUNGMA in your cognitive economy. CHOMOLUNGMA has no substantive a priori entailment relations with your concept MOUNT EVEREST, and you lack sufficient a posteriori reason to think that C1 and C2 corefer.

So, in this case, the antecedent of (5) is met, and we are halfway to a counterexample. How about the consequent? It certainly doesn't intellectually seem metaphysically impossible for their referents to come apart, as (5) predicts. Yet, importantly, neither does it intellectually seem *metaphysically* possible for the referents to come apart. After all, if it really intellectually seemed metaphysically possible for the referents to come apart, then it would intellectually seem that the referents are distinct (since actual identity clearly entails necessary identity). But obviously it doesn't seem so. Nor are you inclined to judge that it is possible for the referents to come apart. You are just agnostic about the identity claim – neither it nor its negation strikes you as obviously true, and you are not inclined to believe either. And so (5) is false no matter how we take the verb "to seem" in the consequent.

Also, (5) overgenerates undercutting defeaters, and this is an additional reason to believe it is false. Nagel's and Hill's proposal would have us be suspicious of some intuitions of distinctness that we know we should not be suspicious of. For example, consider your phenomenal concept of pain and your physiological concept of angiogenesis. (Angiogenesis is the formation and development of blood vessels.) I take it to be obvious that the felt quality type of pain is not (and metaphysically could not be) the formation of blood vessels (that event-type). Clearly and uncontroversially, something could be a token of one type without being a token of the other.

But, if Nagel and Hill are right about phenomenal and physiological concepts, these two concepts have no substantive a priori entailment relations between them. If (5) is true, then I should be suspicious of my belief that the felt quality type of pain is not the formation of blood vessels (that event-type), since, on these suppositions, it would seem metaphysically possible for pain and angiogenesis to come apart whether or not it were metaphysically possible. But clearly I shouldn't be suspicious of that intuition. So this proposal overgenerates undercutting defeaters, and so we shouldn't accept the proposal.

I see no reason in the neighborhood other than (5) to think that actually, our belief in Possibility and Non-Identity is based on

bad grounds. And recall from the discussion of Arnauld, Tye, and Armstrong that it is insufficient to point out merely that *possibly* our belief in Possibility and Non-Identity is based on bad grounds. So I conclude that, as it stands, Nagel and Hill offer us no undercutting defeater.

Furthermore, Nagel and Hill provide us with an undercutting defeater only if the antecedent of (5) is met. But there are good reasons to think that the antecedent of (5) is not met, since there are good reasons to believe that there *are* substantive a priori ties between phenomenal and physical (or physiological) concepts, and specifically between THIS and THAT as deployed in the autocerebroscope case. Consider Non-Identity itself, and some of its cousins. Non-Identity is a substantive a priori entailment relation between THIS and THAT (as deployed during the autocerebroscope case), and the intuition persists even if we swap out the demonstrative concepts for non-demonstrative phenomenal and physical (or physiological) concepts. So there's one reason to believe that the antecedent of (5) isn't met, and therefore that Nagel and Hill do not provide us with an undercutting defeater.

Also, if there are no substantive a priori entailment relations between any two concepts C1 and C2 for some subject S, then it won't seem metaphysically impossible to S for the referents of C1

and C2 to come apart, but it *also* won't seem metaphysically impossible to S that the referents be identical. (Consider CHOMOLUNGMA and MOUNT EVEREST again.) Yet in the dualist thought experiment, it *does* seem metaphysically impossible that the referents of THIS and THAT be identical (since their referents seem actually distinct and actual distinctness clearly entails necessary distinctness).

Finally, if there are no substantive a priori entailment relations between any two concepts C1 and C2 for some normal human subject S, then sufficient empirical information will lead S to believe with no compunction that C1 and C2 corefer. For example, if I gave you reason to believe that CHOMOLUNGMA does refer to Mount Everest, then you would have no problem believing that the relevant concepts corefer. However, no amount of the relevant empirical information closes the explanatory gap in the philosophy of mind – dualist intuitions persist, even in light of all the relevant empirical data. As Papineau (2002, 161) says: "even given all the arguments, intuition continues to object to mind-brain identity." And Tye (1999, 706) says that the explanatory gap remains open "even for those who understand full well the relevant phenomenal terms and who know the underlying physical and functional story." Therefore, again, the antecedent of (5) isn't met, and therefore Nagel and Hill do not provide us with an undercutting defeater.

As it stands, then, Nagel fails to deliver an undercutting defeater. I will now leave Nagel and Hill behind, and move to another type of proposal for an undercutting defeater from Thomas Polger.

5.11. Insufficient Grasp of Relevant Concepts?

Thomas Polger (2004, 49ff) offers a similar account of our dualist intuitions, which he believes are misleading. According to Polger, the culprit is our insufficient grasp of the identity conditions of brain states. Polger thinks that we have, in an important way, failed to fully grasp the concept of a brain state, since we do not know the identity conditions of brain states. And he thinks that, in general, if we are uninformed about the identity conditions of either Xs or Ys, then even if they are identical it might seem that they could come apart. "Thus arises the appearance of contingency," he says. This is a conceptual failure on our part. Our anemic grasp of the concept of a brain state causes us to deem them candidates for identity with phenomenal states, and may cause the relevant identity claim to appear contingent even if it isn't.

This proposal bears a family resemblance to the previous proposal. The idea, I take it, is that our concept of a brain state is really just silent with respect to our phenomenal concepts. Since we

don't fully grasp the identity conditions of brain states, we fail to see that they just are phenomenal states. And, Polger believes, dualists mistake their failure to see that brain states are phenomenal states for successfully seeing that they are not. This idea has something to be said for it, as Polger illustrates with the example of Thingamajigs and Whatchamacallits. If – as suggested by their names – you are unclear on the nature of Thingamajigs and Whatchamacallits, you may mistake the *epistemic* possibility of their distinctness for the *metaphysical* possibility. You may mistake, that is, your inability to rule out their distinctness for your ability to rule it in. This is a tempting slip, at least in the case of Thingamajigs and Whatchamacallits. And perhaps we are making the same mistake when it comes to dualism. If we learn that we are making this mistake, this would successfully undercut our dualist intuitions.

5.12. Response

There are at least two reasons to think that we're not making this mistake when it comes to our dualist intuitions, however. As with the proposal from Nagel et al., Polger's proposal overgenerates undercutting defeaters. We hold many beliefs about brain states that are obvious, uncontroversial, and based on intellectual seemings, and yet which would fall under a cloud of suspicion were Polger correct.

Therefore, Polger is not correct. For example, I take it we all believe that brain states are not numbers, and that brain states are not earthquakes. But according to Polger we are in the dark about the identity conditions of brain states and so we are at risk of mistaking our inability to see the truth of identity claims involving brain states for our ability to see their falsity. It would follow that even a belief as obvious and uncontroversial as that brain states are not earthquakes is undercut, on Polger's view. But since that belief is clearly above reproach, something must have gone wrong with Polger's proposal.

Polger may reply that we are not *completely* in the dark when it comes to the identity conditions of brain states. We know enough about their identity conditions to see that brain states are not earthquakes or numbers, but not enough to see that brain states are not phenomenal states. This is a promising line, but Polger needs to say more to support this claim. How might we distinguish between earthquakes and phenomenal states in a principled way, so it comes out that we know enough about brain states to rule out identity with earthquakes and not phenomenal states? Unless he answers that question, Polger's proposal does not provide us with an undercutting defeater for our dualist intuitions.

Polger's revised proposal might go like this.[21] What we fail to grasp is the fundamental nature of physical matter. We know how matter *functions*, what sort of *relations* it enters into, both at the fundamental level and in aggregate. This is how we know brain states aren't earthquakes: we know that an aggregate of matter that could function as an earthquake could not (or at least would not) function as a brain state. We could then supplement the suggestion that consciousness does not seem to have a functional or relational nature. And it might be that, whenever we entertain two concepts C1 and C2, if C1 is a concept of a thing of which we understand only the function, and C2 is a concept of a thing of which we understand only the intrinsic nature, it might seem that C1 and C2 could fail to corefer even if they couldn't. Finally, one might add that this consequence is sufficient to undercut any intellectual seeming that C1 and C2 could come apart. This would avoid the brain states and earthquakes objection while still promising an undercutting defeater for our dualist intuitions.

However, this proposal still faces difficulty. I take it as obvious that earthquakes are not and could not be phenomenal states. And I take it that if our concept of a brain state is a concept of a thing of which we understand only the function, the same goes with our concept of an earthquake. And so in considering the

[21] I owe this suggestion to David Barnett, though I may not be doing it justice.

identity claim "earthquakes are phenomenal states" I deploy two concepts. One (the concept of an earthquake) is the concept of a thing of which we understand only the function, and the other is a concept of a thing of which we understand only the intrinsic nature. It would follow on the revised proposal, then, that we have an undercutting defeater for our belief that earthquakes are not phenomenal states. This is an unhappy result, and examples could be multiplied. So, this revised proposal overgenerates defeaters and ought to be rejected.

There is a second problem with Polger's proposal, both in the original form and the revised form. Consider again his Thingamajigs and Whatchamacallits example. Since we are clueless about the identity conditions of these things, we may feel a pull toward thinking that they are possibly distinct. But note well that we feel an equally strong pull toward thinking that they are possibly *identical*. If we were to say "Maybe they're identical, but maybe they're not," we'd be reporting epistemic possibilities, i.e. our inability to rule out their identity as well as our ability to rule out their non-identity.

And so, if we are in the same boat with respect to our concept of a brain state, then we should be just as tempted to conclude in dualist thought experiments that brain states are identical with phenomenal states as we are to conclude that they are distinct. They should be candidates for identity as well as candidates for

distinctness. But we are not tempted in that way. Brain states seem like the wrong kind of thing to be conscious states. And so Polger's proposal does not capture the data; it issues a prediction that is falsified by the data. And so I again conclude that Polger fails to provide us with an undercutting defeater for our dualist intuitions. Let's now turn to a final proposed undercutting defeater, from David Papineau.

5.13. Papineau's Proposal

Papineau's strategy for undercutting our dualist intuitions goes like this: call the way in which we think about the phenomenal character of pain from the inside a *phenomenal* concept. (I'll use "$PAIN_P$" to refer to that phenomenal concept.) When we think about conscious experiences in this phenomenal way, when we "deploy" or "exercise" phenomenal concepts, the concepts themselves exemplify or stimulate versions of their respective conscious states, according to Papineau. So when we think about pain in a phenomenal way, when we deploy $PAIN_P$ imaginatively, Papineau (2002, 170) says that "we activate a 'faint copy' of the experience referred to. When we deploy a phenomenal concept introspectively, we amplify the experience referred to into a 'vivid copy' of itself."

Non-phenomenal concepts, on the other hand, do not do this, and so we feel that non-phenomenal concepts "leave out" the feelings themselves. However, the mere fact that non-phenomenal concepts "leave out" these faint or vivid copies does not for a moment suggest that they do not refer to sensations. My non-phenomenal concept C-FIBERS FIRING and my phenomenal concept PAIN_p may still corefer, even though the former "leaves out" the phenomenology that the latter activates. Similarly, the concepts LAUGHING GAS and N_2O may corefer (in fact they do), even if that seems incredible since these concepts activate radically different mental images in us.

According to Papineau, we succumb to the fallacy of thinking that the reason Possibility and Non-Identity seem true has anything to do with the referents of our concepts. Rather, he says, we are committing a use-mention fallacy. A third-person, non-phenomenal way of thinking might not *use* conscious experiences in the way that a first-person, phenomenal way of thinking does. But this fact does not imply that the non-phenomenal concept does not *mention* the same thing that the phenomenal concept does. So although it may seem to us that Possibility and Non-Identity are true, this in no way warrants belief in the truth of Possibility and Non-Identity since the seeming is caused by facts about our concepts, not by facts out there in the world. Possibility and Non-Identity would seem true whether

or not they were true. We are being fooled by a contingent feature of our concepts. And so, if Papineau is right, our dualist intuitions are undercut.

5.14. Response

Let's think about Papineau's argument for the conclusion that we have an undercutting defeater in the case of Possibility and Non-Identity. First, Papineau asserts this (true) proposition:

> (6) For normal human subjects, deployment of non-phenomenal concepts "leaves out" something that deployment of non-phenomenal concepts doesn't.

In addition, Papineau points out the following truth:

> (7) For any concepts C1 and C2, and subject S, the fact that *S's deployment of C1 "leaves out" something that S's deployment of C2 doesn't* does not render probable or warrant belief that *C1 and C2 are not necessarily coextensional (and therefore have distinct referents)* for S.

So Papineau points out a bad basis on which one might judge the truth of Possibility and Non-Identity, namely this contingent feature on our concepts. However, in order to provide an undercutting defeater, Papineau must give us some reason to think that it is not

merely *possible* that Possibility and Non-Identity seem true on this bad basis, but that they *actually* do. Here's an unpromising strategy:

> (8) For any concepts C1 and C2, and normal human subject S, if S's deployment of C1 "leaves out" something that S's deployment of C2 does not, then it will seem true to S on this basis that *C1 and C2 are not necessarily coextensional and have distinct referents*.

If true, (8) would [in combination with (6) and (7)] furnish us withan undercutting defeater, a reason to think that the basis on which we judge Possibility and Non-Identity to be true does not warrant belief in Possibility or Non-Identity.

However, this strategy overgenerates undercutting defeaters, i.e. it gives us reason to doubt the seeming truth of propositions that we rightly take to be indubitable. For example, suppose my friend, hung up on Ockham's razor and seeking to scale down his ontology, proposes that the felt quality type of pain is identical with the felt quality type of euphoria. I consider the identity, deploying the relevant phenomenal concepts. Deployment of each concept "leaves out" something that deployment of the other does not. And it seems obviously true to me that the referents are distinct and not necessarily coextensional. But, if (8) is true, I should be suspicious of this seeming. After all, if, as it says, distinctness of referent and non-necessary coextension will seem true on a bad basis (viz. a contingent

feature of the concepts), then I ought to refrain from judging that things are as they seem. But clearly this judgment is not suspicious – it is as obviously true as anything can be. Therefore (8) overgenerates defeaters, and so it's false. There must be some other basis on which I non-fallaciously judge distinctness and non-necessary coextension.

(8) is also plagued by straightforward counterexamples. Suppose I overhear some friends discussing Smith's favorite color. I don't know what color they're referring to, but I submit to the urge to name it "Colin." I then wonder "Is Colin identical with red?" In considering the identity, I deploy my phenomenal concept of red and my new non-phenomenal concept of Colin. Deployment of the latter concept leaves out something (a "faint copy" of red) that deployment of the former concept does not. And yet it doesn't at all seem to me that the referents are distinct or not necessarily coextensional; I'm agnostic on the question. Therefore, (8) is false, and – as it stands – Papineau fails to provide us with an undercutting defeater.

6. Conclusion

In conclusion, if you share the widespread intuition that Possibility and Non-Identity are true, then your prima facie justification is so far undefeated. You may find fault in the arguments of this paper, or you may know of undercutting defeaters that I do

not discuss and that you take to be compelling. In the following chapter, I will discuss the prospects of an undercutting defeater from the fact that there is widespread disagreement on this issue among epistemic peers. Or, you may have access to a rebutting defeater sufficient to neutralize or override the justification you currently enjoy in favor of dualism. Otherwise, you ought to believe dualism. In the third chapter, I will discuss the prospects for rebutting defeaters.

CHAPTER TWO

The Equal-Weight View Does Not Defeat Dualist Intuitions

Some philosophers believe that when epistemic peers disagree, each has an obligation to accord the other's assessment the same weight as her own. Call this "the Equal-Weight View" of peer disagreement. When it comes to the philosophy of mind, one might think that the quality and quantity of disagreement among philosophers might furnish us with a defeater for the apparent truth of dualism. In this chapter, I will show that this is not so. If a subject believes dualism on the basis of its striking truth in the standard thought experiments, then the Equal-Weight View of peer disagreement will not require the subject to diminish her confidence in dualism.

I first make the antecedent of this Equal-Weight View precise, and then motivate the View by describing cases in which it gives the intuitively correct verdict. Next I introduce some apparent counterexamples – cases of apparent peer disagreement in which, intuitively, one should not give equal weight to the other party's assessment. To defuse these apparent counterexamples, an advocate of the View might try to explain how they are not genuine cases of

peer disagreement. I examine David Christensen's and Adam Elga's explanations and find them wanting. I then offer a novel explanation, which turns on a distinction between knowledge from reports and knowledge from direct acquaintance. Then, I extend my explanation to provide a handy and satisfying response to the charge of self-defeat. Finally, I extend my explanation to the case of dualist intuitions, and show how the Equal-Weight View will not recommend that the dualist give equal weight to the view of a disagreeing epistemic peer.

The Equal-Weight View

Some philosophers believe that when epistemic peers disagree, each has an obligation to accord the other's assessment the same weight as her own. Call this "the Equal-Weight View." Recent advocates include Adam Elga (2007), Richard Feldman (2006), and David Christensen (2007). Elga puts his general view of disagreement this way:

> Your probability in a given disputed claim should equal your prior conditional probability in that claim. Prior to what? Prior to your thinking through the claim, and finding out what your advisor thinks of it. Conditional on what? On whatever you have learned about the circumstances of how you and your advisor have evaluated the claim. (500, n. 26)

The "prior" here needn't be temporal priority. Elga clarifies elsewhere (489-90) that your credence in a disputed claim should equal your conditional probability in that claim *setting aside* "your detailed reasoning (and what you know of your friend's reasoning) about the disputed issue." That is, you are meant to conditionalize on a proper subset of your evidence – a subset which includes what you know of the circumstances of disagreement, but excludes the particular contents of your assessments and any reasoning by which you arrived at them.

Let me now carefully describe a case of peer disagreement, to see what Elga's general view of disagreement recommends. Suppose Smith and Jones disagree about whether *p* on the basis of some shared body of evidence. Smith learns this and believes that (prior to the disagreement) she's as reliable as Jones[22] on the issue given what she's learned about the circumstances of evaluation. Elga's view would say here that Smith's probability in *p* should equal her probability in *p* conditional on these things she's learned, setting aside her and Jones's reasoning on the issue and the content of their assessments themselves.[23] Setting these aside, Elga says, Smith

[22] In this paper, to avoid a cumbersome sentence structure, I intend the admittedly strained reading of "Smith is as reliable as Jones" that entails that both Smith and Jones are reliable. I'm not concerned with cases of disagreement in which both parties are *un*reliable, but to the same degree. I owe that revelation about myself to Nathan King.

[23] For example, if Smith's credence in *p* is 0.8 and Jones's credence is 0.2, the evidence on which Smith conditionalizes should include those facts, but

should think it 50% likely that she's correct (488), i.e., give Jones's assessment the same weight as her own. So, I take it, Elga would agree with the following conditional as an instance of his general view of disagreement:

> **(Equal-Weight View)** For any subjects Smith and Jones and for any *p*, if…
>
> > **(Smith Judges)** Smith's credence in *p* on her evidence E relevant to *p* is *n1*, and
> >
> > **(Jones Judges)** Jones's credence in *p* on E is *n2,* and
> >
> > **(Disagreement)** $n1 \neq n2$, and
> >
> > **(Full Disclosure)** Smith learns these three things, and
> >
> > **(Peerhood)** Smith believes that she's as reliable as Jones on this issue in the circumstances of evaluation, excluding the assessments themselves and any reasoning by which she and Jones arrived at them,[24]

---

should of course exclude *p* itself or any instance of the schema $P(p) = n$.

[24] **Objection:** If Smith excludes Jones's reasoning and her own, then if Smith learns that her reasoning was fantastic and Jones's was shabby, the View will deliver a counterintuitive conciliatory verdict. **Response:** Before Smith's evidence is updated, the View gives the intuitive conciliatory verdict. After the update, if the other antecedent conditions of the View are met, **Peerhood** plausibly won't be, since Smith would sensibly reason in roughly this way: "Jones knows that her assessment (whatever it is) was unreasonable (for whatever reason). Yet nevertheless she sticks with it. So, she suffers from cognitive malfunction and therefore is not as reliable as I am here in the circumstances of evaluation, even setting aside the particular contents of our assessments and any reasoning by which we arrived at them." If Smith doesn't reason this way, then while the antecedent of the View may be met, the conciliatory verdict won't be counterintuitive.

> …then Smith should give Jones's assessment of *p* on this evidence the same weight as her own.

Some philosophers have apparently taken this consequent to entail that Smith's credence in *p* on this evidence should be (at least roughly) the average of *n1* and *n2*.[25] For example, Christensen said[26] that in cases of peer disagreement one should "come close to 'splitting the difference'" (203) between the initial assessments. And – working on an all-or-nothing model of belief and speaking of peers who take equally firm but opposing stances on the disputed issue – Feldman said that after full disclosure, "suspension of judgment is called for." (235) In this paper, I'll be concerned with the Equal-Weight View above, and I won't take a stand on either Elga's general view of disagreement or a general difference-splitting rule for giving equal weight.

Motivating the View

[25] And I suspect some have thought that one's credence in *p* on one's evidence relevant to *p* should equal one's credence in *p* simpliciter, so Smith's credence in *p* simpliciter should also be (at least roughly) the average of *n1* and *n2*.

[26] The past tense in this paragraph is intentional. In light of things they've very recently said and written, I suspect (though I'm not certain) that Christensen and Feldman would no longer say what they then said on this issue.

I take it that many people believe the Equal-Weight View (or something like it) because it delivers intuitively correct verdicts in a wide variety of cases. For example:

> **Feldman's Quad**
>
> Suppose that you and I are standing by the window looking out on the quad. We think we have comparable vision and we know each other to be honest. I seem to see what looks to me like a person in a blue coat in the middle of the quad. (Assume that this is not something odd.) I believe that a person with a blue coat is standing on the quad. Meanwhile, you seem to see nothing of the kind there. You think that no one is standing in the middle of the quad. (223)

In this case, you and Feldman consider all and only the same evidence (namely, the scene before you and any relevant background knowledge). Feldman's visual faculties report to him that there is a person in a blue coat in the middle of the quad; his credence in that proposition on the available evidence is high. Your faculties report otherwise; your credence in that proposition on the evidence is low. And you think your faculties are as reliable as Feldman's. So what should you do in such a case, after full disclosure? Clearly you should revise your belief to give the report of your faculties and the report of his equal weight, just as you would do with disagreeing but equally reliable thermometers, clocks, etc. And so the Equal-Weight View delivers the right result.

The View also gives the right result in some cases involving a priori calculations. For example, Elga (492) and Christensen (193) both consider a case in which friends mentally divide a restaurant check:

> **Restaurant Check**
>
> Suppose that five of us go out to dinner. It's time to pay the check, so the question we're interested in is how much we each owe. We can all see the bill total clearly, we all agree to give a 20% tip, and we further agree to split the whole cost evenly... I do the math in my head and become highly confident that our shares are $43 each. Meanwhile, my friend does the math in her head and becomes highly confident that our shares are $45 each. (Christensen, 193)

To differentiate this case from Feldman's Quad and others crucially involving sense perception, let's stipulate not only that all parties can clearly see the check, but that they all *know* its total. If we stipulate also that all parties think the disagreement is between peers, Christensen and Elga think that, after full disclosure, each should give the other's assessment the same weight as her own. I agree. Here again the View issues the right verdict.

I take it that concrete case intuitions like these strongly motivate the Equal-Weight View. If it weren't for this intuitive support, arguments for the View – such as Elga's Bootstrapping Argument (486-8) – would lose much or all of their force.

Apparent Counterexamples

It's not all sunshine for the Equal-Weight View, however: in some cases it apparently gives the wrong result. Advocates such as Christensen and Elga try to explain why these apparent counterexamples are merely apparent. I'll describe some problematic cases in this section, and in the next I'll lay out Christensen's and Elga's explanations and say why I find them inadequate. Finally, I'll offer my own explanation of these cases, which vindicates the Equal-Weight View.

First, a problematic variation of Restaurant Check:

> **Extreme Restaurant Check**
>
> Consider an (admittedly unrealistic) variant on the restaurant case, in which my friend becomes confident that our shares of the check are $450 – quite a bit over the whole tab.
> (Christensen, 199)

Intuitively, one shouldn't significantly alter her initial assessment of the shares in this case. Christensen and Elga agree. Christensen says, "Here, I think that I need not significantly reduce my confidence in my $43 answer, or raise my very low confidence in the $450 answer." (199) Elga says, "It certainly seems as though you should be more

confident that you are right than that your friend is." (490-1) But, they admit, the Equal-Weight View seems to recommend otherwise.

Consider a new case: **Dual Introspection**. Suppose there's a region of the brain responsible for bodily sensations. And suppose you justifiably believe Alex Byrne when he says "I must have some sort of mechanism… for detecting my own mental states…." (forthcoming) Suppose a trustworthy neuroscientist persuades you that she has hooked up your brain and introspective mechanism with Jones's so that you and Jones now regularly have (at least type-) identical bodily sensations and equal introspective abilities with respect to these experiences.

This neuroscientist causes you (and thereby Jones) to have complicated bodily sensations as of fleeting pains, itches, and tickles, and asks you both to report on the phenomenal character of your experiences. Given your beliefs about the setup and your long track record, you're both comfortable issuing reports of the forms "**We** are experiencing ____," and "**S/he** is experiencing ____," based on introspection. Usually, these reports are true. But due to the kaleidoscopic phenomenology of some of these experiences, occasionally you're mistaken. You learn this. Jones proves to be as reliable as you in her introspective abilities, so you count her as a peer here. You also believe that she's completely honest.

The two of you are asked to introspect a complicated bodily sensation and assess the claim that you (the reader) are experiencing pain. You introspect and find a fleeting sensation that may have been a pain, but then again, perhaps it was just an oppressive itch. Finally, suppose you and Jones disagree about whether you're in pain after full disclosure. It seems that in this case you should give her assessment equal weight. Perhaps she introspectively got a better look at that elusive sensation than you did after all.

However, consider a variant case: **Extreme Dual Introspection**. The setup is the same, except this time you introspect and (seem to) find fierce pain. Your credence in the claim that you (the reader) are experiencing pain is therefore very high. Jones introspects and reports that her credence in this proposition is low. What should you do? Obviously, you shouldn't significantly alter your initial assessment. But the Equal-Weight View seems to recommend otherwise.

An explanation of how its antecedent may not be met in these cases would snatch the Equal-Weight View from the jaws of the apparent counterexamples. Ideally, this explanation would be general enough not to founder on variations of the problematic cases. Christensen and Elga offer such explanations. I'll now describe these and say why I find them unsatisfactory.

Christensen's Explanation

Christensen considers Extreme Restaurant Check and offers this explanation:

> It is much more likely that she calculated and has not brought common-sense checking to bear. Now I take it that this sort of common-sense checking is much less liable to error than mental arithmetic. (201)

Later, Christensen adds:

> The real ground for thinking that my friend made the error in the Extreme Restaurant Case derives from the fact I have evidence that my assessment of the disputed proposition is supported by an extremely reliable kind of reasoning, but I have no basis for supposing the same about my friend's contrary assessment. (201)

The idea seems to be that in this case you come to learn something relevant about the reasoning you and your friend used to answer the question. You learn that she probably didn't use a highly reliable procedure – namely commonsense checking – which you have reason to believe you did use. Conditional on your using commonsense checking and your friend's failing to use it, you shouldn't think that she's as reliable as you on this issue, in which case (Christensen apparently thinks) the **Peerhood** condition in the

antecedent of the Equal-Weight View is not satisfied. If so, the View doesn't issue the counterintuitive conciliatory verdict.

This explanation is unsatisfactory for two reasons. First, a motivating insight behind the Equal-Weight View is that – to avoid being unacceptably arbitrary or question-begging – each party's evaluation of the other as peer or non-peer should be independent of (or "prior to," as Elga says) the content of the disagreeing assessments and any reasoning that led to these assessments. As Christensen says, "I should assess explanations for the disagreement in a way that's independent of my reasoning on the matter under dispute." (199) I find it hard to see how in Extreme Restaurant Check Christensen's evidence that his "assessment of the disputed proposition is supported by an extremely reliable kind of reasoning" is independent of that reasoning.

By way of explanation, Christensen says, "My grounds for discounting my friend's belief are based on considerations about my reasoning, but not on that reasoning itself." (201) But it isn't clear to me that this is so if Christensen's ground for discounting his friend's belief is his evidence that he used commonsense checking. Presumably, Christensen's evidence that he used commonsense

checking is first-personal – based on introspection or memory.[27] It's hard to see what Christensen would be introspecting or remembering, other than his reasoning process itself, in which case his evidence that he used commonsense checking wouldn't be independent of that reasoning. The reasoning hasn't been set aside or bracketed off, it's been remembered or introspected.[28] But if that evidence isn't independent of his reasoning, then according to Christensen's own test, it fails to support an explanation in terms of his friend's failure. So by Christensen's own standard, this cannot be the explanation of why the Equal-Weight View fails to apply in this case.

The second reason Christensen's explanation is unsatisfactory is that it doesn't cover variant cases. Suppose you come to know that in fact your friend did use the same highly reliable procedures you used. In this case, we cannot appeal to any difference in reasoning procedures to justify the claim that **Peerhood** (or any other

---

[27] If that evidence is not first-personal, but based instead on Christensen's track record, dispositions, epistemic virtues, etc., then it's not clear that there is no parallel evidence in favor of thinking that his friend also used commonsense checking. (Or at least there's a problematic case in which there is such evidence.) But in that case we can't accept Christensen's explanation of why **Peerhood** isn't met.

[28] Clearly, Christensen's evidence could be independent of the *result* of his reasoning: his introspective/memorial evidence need not presuppose that his reasoning was *accurate* in the end. I worry that such a narrow principle of independence will license steadfastness when one should be conciliatory, though I can't develop that worry here. **Peerhood** should require setting aside *both* one's answer *and* any reasoning that delivered it.

condition) is not met. And yet even in this variant case, the intuition persists that we should not accord her (obviously false) assessment equal weight. Christensen's explanation does not tell us why this is so.

For instance, stipulate in Extreme Dual Introspection that Jones uses only the same procedure (namely introspection) as you used to arrive at her judgment, so you cannot demote her from peerhood for the reason Christensen suggests in Extreme Restaurant Check. Still, intuitively you shouldn't accord her assessment that you are not in pain the same weight as your own assessment that you are in pain. And so it still seems that the Equal-Weight View delivers counterintuitive verdicts. If the antecedent of the View actually fails to be met in these cases, we don't yet know why.

Elga's Explanation

Elga takes care to point out that **Peerhood** requires conditionalization on the circumstances of disagreement. Elga then says this: "…the circumstances of disagreement might include such factors as: …how absurd each of you finds the other's answer." (490) Later, he adds:

> And one circumstance of the split-the-check disagreement is that you are *extremely* confident that your advisor's answer is wrong – much more confident than you are that your answer is right. Indeed, her answer strikes you as obviously insane. So in order to apply the equal weight view, we must determine your prior probability that you would be right, conditional on these circumstances arising. (491)

Elga's explanation seems to be this: it's natural to assume that there's something about the circumstance of disagreement that should demote your friend from peerhood, namely the fact that you find her answer insane. Assuming that you take the prior probability that you would be right conditional on this asymmetry to be greater than 0.5, **Peerhood** is not met.

However, variant cases are problematic for Elga's explanation. If symmetry is restored – if we stipulate, that is, that your friend also finds *your* answer insane – then, according to Elga, the description of the case doesn't settle whether **Peerhood** is satisfied. Elga considers such a symmetrical case and says that the Equal-Weight View's verdict depends on the answer to the question "Conditional on the two of us disagreeing, and each of us finding the other's answer to be insane, do I think that the two of us are equally likely to be right?" If the answer is "yes," Elga says, then the View rules that you should be conciliatory. Elga says that he finds that plausible.

I, however, find that deeply implausible. And so I don't find Elga's explanation satisfactory. I think that in this symmetrical variant of Extreme Restaurant Check, you should not give your friend's assessment equal weight even if the answer to that last question is "yes," since your friend's answer is obviously false. Given that, you shouldn't significantly alter your assessment, regardless of how she feels about your answer, and regardless of your track record of disagreement with her. You are entitled to believe that your friend's answer is wrong – and therefore not be conciliatory – *come what may*. But Elga disagrees.

For similar reasons, Elga's explanation fares poorly when applied to a variant of the Extreme Dual Introspection case. Recall that Jones introspects and reports that she fails to find any pain. You, however, introspect and (seem to) find fierce, excruciating pain, as though you've stepped in an angry bear trap. Suppose you are both highly confident that the other is wrong, and suppose you have a track record such that you answer "yes" to the question "Conditional on the two of us disagreeing and each of us finding the other's answer insane, do I think that we're equally likely to be right?" According to Elga, you should now significantly weaken – and perhaps even abandon – your belief that you're in pain.

But isn't it obvious that you should not significantly weaken that belief? You are *directly acquainted* with fierce, excruciating pain; the pain is staring (slapping?) you in the face, so to speak. The belief that you're in pain is certain for you, and should not be abandoned no matter what you come to think about Jones's opinions of your belief, your track record, etc. But then Elga's explanation stumbles here. If the Equal-Weight View fails to apply to these cases, we don't yet know why.

My Explanation

I'll now explain why the Equal-Weight View doesn't issue counterintuitive verdicts in the extreme cases described above. If my explanation succeeds in those cases on which Elga's and Christensen's explanations founder, then my explanation should be preferred.

Sometimes we see that p is true by seeing that some other proposition q is true.[29] In those cases, we might say our knowledge that p comes by way of a report, indication, or representation. Other times, we *just* see that p is true, directly. Occasionally – it's said – we just see that p, with our eyes. Here "just see" is used in a literal sense.

[29] For example, the forest ranger sees that the forest floor is on fire by seeing that smoke rises from the treetops.

Looking down, one might say, "I just see that I have hands – there they are, directly in front of me." On other occasions we metaphorically just see that *p*, without the aid of our eyes. For example, we just see that no prime minister is a prime number and that 2 + 2 = 4. It is this metaphorical sense of the ordinary English expression "just see" that interests me for the rest of this paper.

In Extreme Restaurant Check, Smith just sees – in the metaphorical sense – that her friend's answer is wrong. And in the Extreme Dual Introspection case, Smith just sees – in that same sense – that she's in pain. In philosophy-speak, we might say Smith comes to have *knowledge from direct acquaintance* in the problematic cases described above. A relevant piece of evidence is intellectually obvious to Smith; she has unmediated cognitive access to the truth of a pertinent proposition. Her knowledge does not rely on any report, indication, or representation.

And that's why Smith shouldn't be conciliatory about the proposition in question on her evidence: her evidence she can just see – call it "immediately accessible evidence" – includes either her answer or the negation of her friend's answer. Via rational intuition, the proposition that *it's not the case that each share of this check is $450* is part of Smith's immediately accessible evidence in Extreme Restaurant Check. Via introspection, the proposition that *I am in pain*

is part of Smith's immediately accessible evidence in Extreme Dual Introspection.[30] These are pertinent facts about the situations that Smith appreciates; they have thereby entered Smith's cognitive economy.

Given that, in the problematic cases, Smith – sensible person that she is – may reflect on the state of Jones's cognitive economy in roughly the following way: "I just see the truth of a relevant piece of evidence. Jones does as well, or she doesn't. If she doesn't, then I have evidence she lacks, and so **Jones Judges** isn't met. If she does, then either there's merely apparent disagreement,[31] or Jones just sees the truth of some proposition and yet believes it's false. If the former, then **Disagreement** isn't met. If the latter, then here in the circumstances of evaluation, Jones suffers from cognitive malfunction and so is not as reliable as I am on this issue, even setting aside the particular contents of our answers and any reasoning that led us to them."

Full Disclosure and **Peerhood** require that Smith believe certain things. If Smith reasons in this sensible and straightforward way, she'll reject some belief such that at least one of those conditions

[30] N.b., I'm not claiming that a proposition of the form *it seems to me that p* or *I have the intuition that p* gets into one's immediately accessible evidence here. Rather, *p* itself (one's answer or the negation of one's friend's answer) enters one's immediately accessible evidence in these cases.

[31] E.g., Jones is honestly misreporting or dishonestly reporting, or Smith has misunderstood Jones's report.

isn't met. If so, the View's antecedent won't be satisfied in the problematic cases above, and so the View won't issue counterintuitive conciliatory verdicts. (If Smith doesn't reason this sensible and straightforward way, then while the antecedent of the View may be met, the conciliatory recommendation won't be counterintuitive.)

Why does the Equal-Weight View give the intuitive verdict in the non-extreme cases? There, the relevant knowledge Smith gains is *knowledge from reports*: Smith doesn't just see that *p* but rather receives a report that *p* from her faculties. Smith then learns of the disagreeing report of Jones's faculties. And if someone has disagreeing reports from two sources she takes to be equally reliable, then *ceteris paribus* she should give them equal weight. This goes for thermometers, clocks, and – in the case of Feldman's Quad – visual faculties.

We non-savants enlist a cadre of cognitive faculties for complex calculations, and we lean heavily on memory. In Restaurant Check, we don't just see that each share of the check is $43 (though we do just see that each share of the check isn't $450). Rather, our faculties report the answer, after some complicated calculations. And if I take my faculties to be as reliable as yours, then I should be conciliatory when I learn of the disagreeing report from your faculties.

In the non-extreme Dual Introspection case, though I am directly acquainted with pain (if it's there), its elusiveness prevents a level of attention sufficient to come to know from this acquaintance that *I am in pain*. While I just see the pain (if it's there), I don't just see that *I am in pain*. I judge that I am in pain, but here my judgment crucially relies on the reports of memory. When I receive the report of your faculties in the case as described, I should be conciliatory.

Rational intuition and introspection do not merely give us knowledge from reports. With at least some cases of introspection and rational intuition, there is no appearance/reality distinction, and so no appearance that reports or represents reality. Introspection does not merely *represent* that there is pain, as my visual experience represents that there is a computer before me. No, by introspecting I can become directly aware of pain itself, and with sufficient attention I can thereby just see that I am in pain. Rational intuition does not merely testify that 2 + 2 = 4 as my kindergarten teacher did. No, via rational intuition I can just see that 2 + 2 = 4. And while it would be unacceptably arbitrary to dismiss the report that *p* from your friend's faculties on the basis of the report that not-*p* from your own faculties (when you take your faculties to be equally reliable), it is not unacceptably arbitrary to do so on the basis of not-*p* when you just see that not-*p*. In fact, such steadfastness is called for.

Our ability to just see the truth of propositions distinguishes us from thermometers, clocks, etc., which merely report, and explains why the Equal-Weight View does not issue counterintuitive recommendations. By giving such an explanation, I have vindicated the View from those apparent counterexamples. And my explanation covers variations of these cases on which Christensen's and Elga's explanations founder. Therefore, my explanation should be preferred. Let me now vindicate the View from one more objection.

Self-Defeat?

Critics have charged that if an adherent of the Equal-Weight View knows of even one equally informed peer who disbelieves it strongly enough, then giving the peer's assessment equal weight will require giving up the View itself. The critics often graciously volunteer to play the role of the disagreeing peer. So – they conclude – if the View is true, we shouldn't believe it. And of course if it's false we shouldn't believe it either.[32]

But this objection should not trouble the adherent of the Equal-Weight View, since the explanation I gave above provides a handy and satisfying response. There we learned how the antecedent of the

[32] Plantinga (2000) offers this type of objection.

Equal-Weight View might not be satisfied in cases involving knowledge from that unmediated access to the truth of propositions sometimes afforded by rational intuition. And it's plausible that the Equal-Weight View is itself a deliverance of rational intuition. Even Thomas Kelly, a prominent opponent of the View, admits that "reflection on certain kinds of cases can make it seem almost trivial or obviously true." (forthcoming)

With further reflection, I think, one can come to just see the truth of the View – not only does it *seem* obvious, but upon further reflection it just *is* obvious.[33, 34] Its non-adherents have, for all their virtues, failed to fully appreciate this. And if an adherent of the View does just see its truth, its antecedent will not be satisfied when she reflects on the skeptic's cognitive economy in the way described above. If so, the View won't recommend giving itself up merely because there are intelligent, informed, and firm disbelievers of the View, and so the View won't be self-defeating.

---

[33] Though not, of course, *as* obvious as, e.g., that *each share of this check isn't $450* in Extreme Restaurant Check. Obviousness comes in degrees.

[34] **Objection**: The Equal-Weight View is complicated and obscure, and so not plausibly a proposition one can just see the truth of. **Response**: Don't sell yourself short. Also, it often happens that a complicated and obscure *sentence* expresses an obviously true *proposition*. For many of us, this is the case with "Kein Premierminister ist eine Primzahl." Likewise with some statements of the Equal-Weight View, I believe.

Dualist Intuitions

Suppose the Equal-Weight View is true. Now consider a subject who believes dualism on the basis of an intellectual seeming. In the standard thought experiments, she claims to *just see* that dualism is true. One might suspect that, if this subject knows of an equally informed peer who disbelieves dualism strongly enough, then giving the peer's assessment equal weight will require giving up the View itself. And so, one might suspect that we have here a defeater for our dualist intuitions.

But these considerations do not furnish us with a defeater for our dualist intuitions, even if we adhere to the Equal-Weight View. Above, we learned how the antecedent of the Equal-Weight View might not be satisfied in cases involving knowledge from that unmediated access to the truth of propositions sometimes afforded by introspection and rational intuition. And it may well be that, in the standard thought experiments, one comes to have knowledge of the truth of dualism via exactly this sort of direct acquaintance. As we noted in the first chapter, even committed non-dualists report the intuitive appeal of dualism.

Therefore, if an adherent of dualism does just see its truth, the antecedent of the Equal-Weight View will not be satisfied when she reflects on the non-dualist's cognitive economy in the way described

above. If so, the View won't recommend giving up dualism merely because there are intelligent, informed, and firm disbelievers of dualism, and so the Equal-Weight View does not provide a defeater for dualism held on the basis of intuition.

Objection

One might object this way: "Intellectual seemings are defeasible. Or at the very least it is illicit to assume that they are indefeasible for the purposes of the Methdological Argument of this dissertation. We have been looking for undercutting defeaters of our dualist intuitions, after all. And so the only evidence that we have a right to assert is that Possibility and Non-Identity *seem* true. It's illegitimate to assert that they just *are* true, i.e. that those propositions are part of our immediately accessible evidence, suitable for playing the role I put them to in the face of apparent peer disagreement.

But then consider a genuine case of disagreement between the dualist and the non-dualist, both of whom share these intellectual seemings. Surely such disagreements actually happen. (Though Possibility and Non-Identity seem true to the non-dualist, she ultimately rejects these seemings as misleading in light of her total evidence.) In such a case there is no asymmetry that the dualist can point to in order to evade the antecedent of the Equal-Weight View.

And it's implausible that the non-dualist is suffering from cognitive dysfunction merely for rejecting an intellectual *seeming*. But then, supposing that the Peerhood condition could be met (and it no doubt could), the antecedent Equal-Weight View will be satisfied, and it will recommend that the dualist give up her belief. And thus we have an undercutting defeater for dualism."[35]

Response

What Extreme Restaurant Check and Extreme Dual Introspection show, I believe, is that it is sometimes perfectly legitimate for a party to a disagreement to recognize that the content of her belief occupies a special place in her evidence base. Though it would be illegitimate to reason from the content of the disputed claim to the conclusion that the other party to the debate is in error, these cases seem to show that it is acceptable for the subject to admit that the contested claim (or one that straightforwardly entails that her friend is wrong) is part of her immediately accessible evidence. This, I think, explains why we need not be conciliatory in cases of extreme disagreement.

If the objector is right, it's hard to see why we shouldn't be conciliatory even in cases of extreme disagreement. But that's quite counterintuitive. For suppose that the objector is right and all that a

[35] I owe this suggestion to Adam Pautz, though I may not be doing it justice.

subject to an extreme disagreement can legitimately recognize as part of her evidence is that it *seems* to her that p. And suppose the objector is right that it's possible for p to seem true to a subject, and yet nevertheless the subject fails to believe p in full epistemic propriety. And suppose there are extreme disagreements in which the contested proposition intellectually seems true to both parties. We can stipulate that this occurs in Extreme Restaurant Check, for example.

If the objector is right, and if the Equal-Weight View is true, then it's difficult to see why the Peerhood condition would fail to be met. But then we ought to be conciliatory in this version of Extreme Restaurant Check. For what it's worth, I think that is a disastrous result. Your friend tells you each share of this check is $450. This strikes you as obviously false. In fact, it intellectually seems false to your friend as well, but only in a sense of "intellectually seems that p" that is consistent with rejecting p while not suffering from cognitive dysfunction. Must you really reduce your confidence that each share of this check is not $450? I think not. And so I conclude that either the objector is wrong or the Equal-Weight View is false. That is, either it is perfectly legitimate for a subject to a disagreement to recognize that the propositions delivered to her by intuition are part of her evidence and so reason in the way described above in *My Explanation*, or the Equal-Weight View is false. Either way, the dualist will be off the hook when it comes to disagreement with well-

informed non-dualists, even those who share her dualist intuitions. Either the objector is wrong and so the story I told above is true: the Equal-Weight View is silent in such a case. Or, the Equal-Weight View is false and so it cannot issue binding conciliatory verdicts on the dualist.

CHAPTER THREE

Rebutting Dualism

So far, we have learned that there are no successful undercutting defeaters for our dualist intuitions. We have also learned that conciliatory views of peer disagreement would not provide a defeater for dualist intuitions. In this chapter, we will investigate the prospects for a rebutting defeater for our dualist intuitions.

Definitions

Let's begin with a few definitions. For any proposition *p*, a *defeater* of a subject Smith's belief that *p* is some evidence E for another proposition *p** such that, if Smith believes that *p* and then gains access to E and takes it *as* evidence in favor of *p**, then Smith must believe that *p* less strongly.[36] Something counts as a *rebutting* defeater for Smith's belief that *p* iff it is a defeater, and the defeating evidence involved is evidence for the proposition that *not-p*.

[36] Here, I do not mean a merely psychological notion, i.e. that as a matter of nomic necessity (given one's physical and psychological make-up), one cannot persist in believing that *p* as strongly as before. Rather, I mean that, given certain epistemic norms, one shouldn't persist in believing that *p* as strongly as before. One cannot so persist in full epistemic propriety.

Let's say a rebutting defeater for Smith's belief that *p* is merely *diminishing* iff when Smith comes to accept the defeating evidence *as* evidence for *not-p*, she can reasonably persist in believing that *p*, though not as strongly as before. Let's say a rebutting defeater for Smith's belief that *p* is *neutralizing* iff when Smith comes to accept the defeating evidence *as* evidence for *not-p*, she can't reasonably persist in believing that *p*, but also can't reasonably believe that *not-p*; Smith must be agnostic about the issue. Finally let's say a rebutting defeater for Smith's belief that *p* is *overriding* iff when Smith comes to accept the defeating evidence *as* evidence for *not-p*, she can't reasonably persist in believing that *p* and must now believe that *not-p*.

Standards for Success

Not just any evidence is sufficient for an overriding rebutting defeater. The strength of a rebutting defeater is a function of the type of defeating evidence. Say I believe that *p* on some evidence E: what it is reasonable or sensible to do in light of new evidence E* for *not-p* will depend on which evidence – E or E* – is "better" or "more credible" in some intuitive sense.

For example, say that I take Alan to be a generally reliable source of information. Alan testifies that he saw Bob buy a Frisbee at noon, I come to believe on this basis that Bob bought a Frisbee at

noon (call this proposition *p*). But suppose that later, I receive testimony from Bob – whom I take to be as generally reliable as Alan—that he was sick in bed all day, and has never bought any Frisbees at all. In this case, my evidence in the form of Bob's testimony overrides my evidence for *p*, and I should come to believe that *not-p*. Since I take it that Bob was in a better position to know whether *p*, and since he testifies that *not-p,* Bob's report supplies me with an overriding rebutting defeater for my belief that *p*. The better and more credible evidence from Bob trumped the evidence from Alan.

Now suppose that Alan and Bob were in an equal position to know the truth of some proposition, say the content of the mayor's speech last week. I take them to be equally honest, to have equal auditory, vocal, and memory capacities, and to have both been in an equally suitable position to hear the mayor's speech. And suppose these beliefs of mine are all true. Alan testifies on the basis of memory that the mayor admitted to corruption (call this proposition *q*). Having no reason to think otherwise, naturally I come to believe *q* on the basis of this testimony. But just then Bob testifies on the basis of memory that it's not the case that the mayor admitted to corruption (*not-q*).

In this case, the testimony from Alan that *q* and testimony from Bob that *not-q* cancel out, so I withhold belief in *q* and I withhold belief in *not-q*. In the absence of any other evidence, I should be agnostic on the question of whether the mayor admitted to corruption. Here, Bob's report has supplied me with a neutralizing rebutting defeater. There are better and worse sources of evidence, and when I have conflicting evidence, I favor what I take to be the better source. Only when I take the sources to be on a par should I suspend judgment – only then should I be agnostic on the issue.

Finally, suppose that Alan, a brilliant mathematician, testifies to me that 7+5=75. On the basis of intuition, I believe that 7+5=12. Is the evidence of Alan's testimony so credible as to furnish me with an overriding or neutralizing rebutting defeater? I believe it is not. At *most*, this evidence is a diminishing defeater. The intuitive basis of my belief that 7+5=12 is better than the evidence I have in favor of the proposition that 7+5=75.

So it goes with expert testimony versus intuition, but what do we say about the case of powerful but abstruse philosophical arguments? What sort of rebutting defeaters can they offer to strong intuitions? For example, consider this argument for the conclusion that 2=1: First, let $a=b$, where a and b are non-zero quantities. Then, multiply both sides by a, to get $a^2=ab$. Now subtract both sides by b^2,

to get $a^2-b^2=ab-b^2$. Next, factor both sides to get $(a-b)(a+b)=b(a-b)$. Then, divide both sides by $(a-b)$ to get $(a+b)=b$. Since we were given that $a=b$, we can infer that $b+b=b$. By combining like terms on the left, we get that $2b=b$. Finally, if we divide both sides by the non-zero b, we get the conclusion that 2=1, Q.E.D.

Does this argument furnish us with an overriding or neutralizing rebutting defeater for our intuitive belief that $2\neq1$? Clearly not. In general, deductive arguments are merely invitations to compare subjective probabilities. On the one hand, we consider the probability of the conjunction of the argument's premises together with each of the argument's inferences in the form of conditionals. On the other hand, we consider the negation of the conclusion. Together, this conjunction and the negation of the conclusion form an inconsistent set. So, one cannot rationally believe both members of the set: at least one must go.

And so, if the negation of the conclusion of this argument deserves sufficiently higher credence than the conjunction of the premises or inferences, then the right thing to do is to reject the conjunction. In the argument of the previous paragraph, the conjunction of all the steps is far more dubious than the proposition that $2\neq1$. Therefore, the appropriate thing to do here is to persist in

the belief that 2≠1 and to suspect some catch in the argument, *even if one cannot identify the catch*.

Rebutting Defeaters for Dualism

A rebutting defeater for dualism would consist of an argument for non-dualism or an objection to dualism. In this section, I will lay out the most common examples of each type of rebutting defeater. I will refrain from responding specifically to these arguments. Instead, I will describe in the next section a general strategy for responding to these defeaters, using David Papineau's Causal Argument as a paradigm example.

First, arguments for non-dualism (cf. Stoljar 2009). Many arguments for non-dualism rely on some sort of causal exclusion principle. These arguments typically proceed by first assuming that every event that has a cause has a physical cause. The second premise is that mental events have physical effects. These arguments then assert some variation of a causal exclusion principle (see for example Kim 1993, Melnyk 1994, Peacocke 1979, and Yablo 1992). While the details vary, the principle asserts roughly that for any cause C with effect E, there is no cause C* of E which is not supervenient on C. Conclusion: mental events are supervenient on physical events. While such a supervenience claim is not strictly

incompatible with dualism as I have defined it—since at least one mental type may not be identical with any naturalistically-acceptable type even while the mental supervenes on the physical—it is inconsistent with our dualist intuition of Possibility.

A second line of argument in favor of non-dualism is inductive. For example, J.J.C. Smart (1959, 142) reasons as follows:

> [S]cience is increasingly giving us a viewpoint whereby organisms are able to be seen as physico-chemical mechanisms. . . . There does seem to be, so far as science is concerned, nothing in the world but increasingly complex arrangements of physical constituents. All except for one place: in consciousness. . . . That everything should be explicable in terms of physics . . . except the occurrence of sensations seems to me to be frankly unbelievable.

The argument, I take it, proceeds by enumerative induction. Smart begins by claiming that a wide variety of phenomena lend themselves to wholly physical explanation. The conclusion generalizes to all varieties of phenomena, claiming that everything—including the mental—should be explicable in terms of physics. Again, while this conclusion is not by itself inconsistent with dualism as I have defined it, it is plausible that it is inconsistent with our dualist intuition of Possibility. If Possibility were true—if mental types could be instantiated without any candidate naturalistically-acceptable type or vice versa—it's hard to see how the mental could

be wholly explicable in naturalistic terms. (See also Melnyk 2003 for a sustained defense of this type of inductive argument.)

Smart also offers a related argument, which appeals to an alleged theoretical simplicity of non-dualism vis-à-vis dualism. He says (ibid., 142-3) that if dualism were true sensations would have to be "nomological danglers," and that "it is not often realized how odd would be the laws whereby these nomological danglers would dangle… I cannot believe that ultimate laws of nature could relate simple constituents to configurations consisting of perhaps billions of neurons… Such ultimate laws would be like nothing so far known in science." The idea, I take it, is that dualism posits an unnecessary number of fundamental types and laws, and that the laws connecting these types to the constituents of brains would be enormously complex. For these two reasons, then, simplicity favors non-dualism.

A fourth and final type of argument for non-dualism appeals to methodological naturalism. Here's how Stoljar (2009, §16) puts it:

> The first premise of this argument is that it is rational to be guided in one's metaphysical commitments by the methods of natural science. Lying behind this premise are the arguments of Quine and others that metaphysics should not be approached in a way that is distinct from the sciences but should rather be thought of as continuous with it. The second

> premise of the argument is that, as a matter of fact, the metaphysical picture of the world that one is led to by the methods of natural science is physicalism. The conclusion is that it is rational to believe physicalism, or, more briefly that physicalism is true.

For this argument to cut against dualism as I have defined it, one must take the second premise to assert that the metaphysical picture of the world best supported by the natural sciences is one in which each mental property type is identical with some naturalistically-acceptable property type.

This concludes, I believe, a survey of the most influential and plausible arguments for non-dualism. Before turning to a general response to these arguments, I will first describe another type of rebutting defeater for dualism: objections to dualism.

One popular type of objection to dualism is Kim's Pairing Problem (see Kim 2005, 70–92). The objection goes like this. First, Kim argues that causal interaction between objects requires that those objects be spatially situated. Next, Kim points out that, on standard forms of Cartesian dualism, souls are not spatially located. Conclusion: souls don't cause anything to happen. Many would consider this a cost of dualism, if the argument is sound.

Dean Zimmerman (2010) argues that a mere dualism of property types (what I've been calling "dualism") without a corresponding dualism of the substances that bear those properties (what is commonly called "substance dualism") is problematic. First, Zimmerman argues against "act-object" theories of perception, on which phenomenal properties like red are properties of objects. On these theories, phenomenal redness would be a property of sense data, brains themselves, or external objects. The first two are implausible, according to Zimmerman, and the third option is contrary to the spirit of property dualism. So, Zimmerman concludes, the property dualist ought to accept an adverbial or intentionalist account of phenomenal properties.

The problem for the property dualist arises, according to Zimmerman, if she identifies the subject of experience with a "garden-variety" physical object, such as a brain or body. For these objects are vague in their spatial and temporal boundaries. And "[g]iven what we know about the close connections between brain activity and phenomenal experience in our own case, laws of qualia generation have, *very* roughly, the form: whenever some neurons are organized and behaving like *so* — e.g., like the ones in my brain right now — something-or-other will be caused to have such-and-such fundamental phenomenal property." But what is this "something-or-other"? Brains and bodies have vague boundaries—there are many

eligible candidates for being my brain or my body. So the property dualist who identifies the subjects of conscious experience with garden-variety material objects has a trilemma: either the laws of nature governing qualia generation are prodigal, exact, or under-generative.

Prodigality: perhaps the laws of qualia generation target all the precise candidates for what we mean by "brain" or "body" *and more*. If I am the subject of experience, then, "I" will not pick out a garden-variety material object. Exactness: perhaps the laws of qualia generation target all and only the precise candidates for what we mean by "brain" or "body." But this would implausibly "attribute to nature itself a touching deference to our linguistic practices and to our rough-and-ready concepts." Under-generation: perhaps the laws of qualia generation target some proper subset of the precise candidates for what we mean by "brain" or "body." Then again I would fail to be a garden-variety material object, and the brain or body "will be at best *sort of* conscious. Whatever else I know about myself right now, I know that I am *definitely* conscious; so if a smaller thing or things *definitely* have the adverbial qualia, I am not the thing that is only indefinitely conscious; I am that smaller thing…"

Thus, Zimmerman concludes, the mere property dualist is pushed into a sort of "speculative materialism," on which subjects of

conscious experience are not garden-variety material objects but rather bizarre metaphysical entities. This, Zimmerman concludes, warrants the conclusion that substance dualism is a live option and in many ways preferable to mere property dualism. Insofar as he is correct, this is an objection to mere property dualism.

Robert Adams (1987, 243-62) argues that mere dualism unconjoined with theism is untenable. The argument goes as follows: A scientific explanation of a correlation is adequate only if (i) the law in terms of which it is explained must be more general than the correlation, i.e. the law correlates things that do or could occur more widely than the terms of the correlation to be explained, and (ii) the explanation does not presuppose any of the facts to be explained. Adams then argues that, absent a theistic explanation, the requisite generality is impossible in the case of mind-body correlations. The reason is that in order to explain the correlation between a brain state B and sensation S, one would have to find a physical state the description of which (in more general terms) B uniquely satisfies and is correlated with a conscious state the description of which (in more general terms) is uniquely satisfied by S. There are no such general descriptions, Adams argues, because such general descriptions would require analyzing sensations as structured complexes (but some are simple not all are amenable to such analysis), or arranging sensations on a scale, assigning numerical values to them, and

discovering an algorithm for finding the numerical value of the corresponding sensation given a numerical value determined by certain quantities in a physical state. But this too is impossible, since sensations do not admit of such an arrangement over a range of mathematical values (in any non-ad hoc or non-question-begging way). Even if we could arrange the sounds in a sound space ordered on pitch, loundness, etc. and even if we could arrange the colors in a three dimensional color space (hue, brightness, and saturation), "[t]he chief difficulty with this strategy is that these orderings cannot be extended to the other sensory modalities, and are not naturally integrated with each other" (1987, 257). Thus, Adams concludes, dualism without theism leaves mind-body correlations unexplained. And that, no doubt, is a cost for any view of the mind.

This completes a brief survey of some of the most prominent rebutting defeaters for dualism that have been proposed. It would no doubt require much space to respond to each in detail. Instead, I will provide a general strategy for responding to such rebutting defeaters, using one paradigmatic example. What I say about the paradigmatic example will apply to any proposed rebutting defeater for dualism.

A General Response to Rebutting Defeaters

I will now lay out a general response to rebutting defeaters for dualism, using Papineau's (2002, 17-18) Causal Argument as a paradigm example:

> (1) Conscious mental occurrences have physical effects.
>
> (2) All physical effects are fully caused by purely physical histories.
>
> (3) The physical effects of conscious causes aren't always overdetermined by distinct causes.

Therefore,

> (4) Materialism is true[37]

Clearly, if I do find it persuasive, such an argument for (4) gives me reason to believe that (4) is true. But how does this kind of evidence compare to intellectual seemings, for example the intuitive truth of Possibility and Non-Identity? Papineau's argument concludes that (4) is true. Possibility and Non-Identity seem true, and either are equivalent to or obviously entail the denial of (4). Would Papineau's argument, granting that we find it persuasive, provide an overriding or even a neutralizing rebutting defeater for our belief that (4) is false?

I believe it would not, for the same reason we should maintain belief that 2≠1 despite the argument to the contrary above.

[37] Papineau uses "materialism" broadly to mean that phenomenal types just are physical or functional types.

If I judge the conjunction of the premises and inferences of Papineau's argument to have a sufficiently lower probability than the denial of the conclusion, then the argument gives me neither an overriding nor a neutralizing rebutting defeater. And I think it's clear that the conjunction of the premises and inferences of Papineau's argument deserves less credence than Possibility and Non-Identity.

Consider, for example, the main inferences of Papineau's argument. Papineau believes that his premises entail his conclusion. After laying out the premises, he says: "Materialism now follows." However, Papineau's argument is not valid. To show this, I will describe a coherent view on which Papineau's premises are true but his conclusion is false. I will first describe the view, and then I'll explain why it isn't a form of materialism.

Here is the view. Suppose all sensations are irreducibly non-material (in Papineau's sense of "material"). At the same time, suppose that beliefs are completely reducible to material properties . And so beliefs supervene on the material facts with at least metaphysical necessity. All the sensations, on the other hand, supervene on their corresponding physical states with at most nomic necessity.

Now let's flesh out the view so that it satisfies the Causal Argument's premises. Add to the view that conscious mental

occurrences have physical effects (so the first premise is true). Add also that all physical effects have prior sufficient physical causes (so the second premise is true). Therefore, on this view we are considering, *much* of the time the physical effects of conscious causes are overdetermined. For example, when the thrill of victory causes me to smile, my smile will have two prior sufficient causes on this view: the irreducibly immaterial thrill sensation and also its purely material correlate. But the physical effects of conscious causes are *not always* overdetermined. For example, when the belief that the Causal Argument is invalid causes a materialist to frown, there is no overdetermination.[38] On this view, the belief just is part of the physical history of the materialist's frown. And so the third premise is true on the view I am describing.

So, all the premises of the Causal Argument come out as true on the view I am describing. Now consider the conclusion. Would materialism be true on this view I have described? Insofar as we can have intuitions about semi-technical terms like "dualism" and "materialism," I should think it sufficiently clear that this view is *not* a form of materialism: every single sensation is irreducibly

[38] If for some reason you think that occurrent beliefs do not count as conscious mental states, then change the view I describe thusly: nearly every sensation is irreducibly immaterial, except for one. The taste of banana, say, is reducible. It just is a brain state. When the taste of banana causes a subject to say "Yum!," there is no overdetermination. In every other case of a physical effect of a conscious cause, however, there will be overdetermination. And so all three premises of Papineau's Causal Argument can be true on this view.

immaterial. But we need not rest our case on such intuitions. Though it is difficult to say exactly what materialism is, many philosophers agree that any adequate definition should at least entail this modal supervenience claim: the mental supervenes on the material with something stronger than nomic necessity.

That supervenience claim is false on the view I have described, and therefore this view is not a form of materialism. There are materially identical possible worlds that nevertheless differ with respect to mental facts. To be sure, on this view the facts about beliefs cannot vary unless the material facts vary. But that does not hold with respect to *all* mental states on this view. For example, on the view I have described it is perfectly possible to have two worlds that are duplicates with respect to the material facts and yet which vary with respect to the facts about sensations.

Therefore, the premises of this Causal Argument could be true even while the conclusion is false, and so the argument is invalid. Now, the argument would be valid were we to strengthen the third premise to something like this: the physical effects of conscious causes are *never* overdetermined by distinct causes. But that is a substantially stronger claim than Papineau's original premise, and no argument is given for this stronger version of the third premise. And while one might easily be skeptical of rampant

overdetermination, it is more difficult to get worked up about the suggestion that it *sometimes* happens. Indeed, many philosophers already accept that *some* physical effects are overdetermined. Why not think that the physical effects of conscious causes are occasionally overdetermined? Therefore, the stronger version of the third premise has much less going for it. By itself, the revised third premise is less plausible than the negation of the conclusion. For even stronger reasons, the conjunction of all the premises and inferences in Papineau's revised argument will be less plausible than the negation of the conclusion. And so, even Papineau's revised argument will fail to deliver an overriding or even neutralizing rebutting defeater for dualism.

Even worse, the premises of Papineau's causal argument are inconsistent with role functionalism and therefore, from the perspective of the contemporary materialist, with multiple realizability. Thus we have even stronger reasons to reject the conjunction of Papineau's premises. What follows has the structure of a dilemma. Papineau considers two natural readings of the first premise of the Causal Argument. On the first reading, I will argue, role functionalism is inconsistent with the first premise. On the second reading, I will argue, role functionalism is inconsistent with the third premise. Either way, then, role functionalism is inconsistent

with this Causal Argument's premises. I will end by explaining why this is a serious strike against the argument.

According to role functionalism, pain is a second-order state defined by its functional role. Roughly, it is the state of being in some state or other that is typically caused by bodily injury, which typically causes a desire for relief, anxiety, evasive behavior, etc. This "some state or other" is often called the *realizer* of the second-order state. In humans, the realizer of pain is C-fibers firing (or whatever). In octopi, the realizer is something else. Both humans and octopi feel pain, according to role functionalism, because each has some state or other that realizes the appropriate second-order functional state.

Papineau (following Malcolm 1968 and Kim 1989, 1998) expresses concern that role functionalism is inconsistent with the first premise of his argument. After all, suppose we kick a human in the shin and he subsequently limps. "I am limping because my leg hurts," he sincerely says. But could that be true, on role functionalism? Is it really the pain—that *second-order* state—that causes him to limp? Or is it the first-order physical realizer of the pain that causes him to limp? Papineau is initially inclined toward the latter. In a strict sense of "cause," he says, only the first-order state causes the limp. The second-order functional state—i.e. the pain itself, on role functionalism—does not cause the limp. But according

to role functionalism all mental states are second-order states. And so what goes for pain goes for every other mental state, on role functionalism. And therefore conscious mental occurrences *never* have any physical effects, on role functionalism. But that is contrary to the first premise. Therefore, a natural reading of the first premise is inconsistent with role functionalism. This is the first horn of the dilemma for Papineau.

However, there are other readings of the first premise. Some of them are not unnatural. And some of these not unnatural readings are consistent with role functionalism. Papineau (2002, 32) says, for example, that "it is arguable that there is a perfectly normal sense of 'cause' in which higher states cause the effects that their realizers cause." And so he draws our attention to the following reading of the first premise:

> (1*) Conscious causes have physical effects, at least in the generous sense.

Role functionalism is consistent with (1*). But I will now argue that, on this reading of the first premise, role functionalism is inconsistent with the third premise of the Causal Argument, i.e. the premise that the physical effects of conscious causes are not always overdetermined by distinct causes. This is the second horn of the dilemma for Papineau.

On this generous sense of 'cause', it turns out that pain might be a second-order functional state and still cause physical effects. As Papineau (2002, 32) says, "my taking an aspirin can still be caused by the pain in my head, in virtue of being caused by whichever strictly physical state realizes that pain in me." But then there are *two* nomically sufficient causes of the aspirin-taking. And, one might think, if an event E has two distinct nomically sufficient causes C1 and C2, then E is overdetermined. It would follow, then, that my aspirin taking is overdetermined. And what goes for pain and aspirin-taking seems to hold generally. Therefore, it looks as though the physical effects of conscious causes will always be overdetermined on role functionalism given this generous sense of 'cause'. But that is inconsistent with the third premise of the main argument. Therefore, even on this more generous reading of the first premise, the Causal Argument is inconsistent with role functionalism. That concludes the second horn of the dilemma for Papineau. Let me now defend the second horn from an objection from Papineau.

Papineau is aware of the worry that constitutes the second horn of the dilemma. In response, he would deny the above principle that *if an event E has two distinct nomically sufficient causes C1 and C2, then E is overdetermined.* He would insist that merely having two nomically sufficient causes is insufficient for overdetermination,

since it is also necessary that these causes satisfy certain counterfactuals. Here is what he says (2002, 33):

> The higher cause is present only in virtue of the physical cause which realizes it. In the circumstances, the one would be absent if the other were. And because of this, we have no trouble with the counterfactuals which would be indicative of genuine overdetermination. It is not true that the behavioural result would still have been caused even if the physical realizer had been absent, for the higher state would then have been absent too; and similarly, if the higher state had been absent in some particular case, there would again have been no alternative cause for the behavioural result, since the physical realizer would have had to be absent too.

I will now reconstruct Papineau's argument for the conclusion that (1*)-(3) are consistent with role functionalism. I will then show how the argument fails. Thus will the second horn of the dilemma for Papineau remain compelling.

Consider again the case in which my taking an aspirin is caused by the pain in my head *in virtue of* (as Papineau says) being caused by whichever strictly physical state realizes that pain in me. According to Papineau

> (A) If the physical realizer had not occurred, the pain would not have occurred.

And therefore, says Papineau,

(B) It is false that I still would have taken aspirin even if the physical realizer had not occurred.

Similarly, Papineau says,

(C) If the pain had not occurred, the physical realizer would also not have occurred.

And therefore, says Papineau,

(D) It is false that I still would have taken the aspirin even if the pain had not occurred.

Putting (B) and (D) together, we get

(E) It is false that (**EITHER** *I still would have taken the aspirin even if the pain had been absent* **OR** *I still would have taken aspirin even if the physical realizer had been absent*).

And from (E) we are meant to draw the final conclusion that

(F) My taking the aspirin is not overdetermined by my pain and the physical realizer of my pain.

In the move from (E) to (F), Papineau assumes the following universal principle:

(G) Any effect E is overdetermined by any causes C1 and C2 only if (**EITHER** *E still would have occurred even if C1 had not occurred* **OR** *E still would have occurred even if C2 had not occurred*)

That is, Papineau places a necessary condition on overdetermination. This condition may not be met even if an effect has two distinct nomically sufficient causes. On role functionalism, in the case of physical effects of conscious causes, this condition is not met. This is why Papineau believes that role functionalism is consistent with premises (1*)-(3).

Here is my criticism of this argument: principle (G) is false. The condition Papineau takes to be necessary for overdetermination is not necessary. To show that principle (G) is false, I will describe a clear instance of overdetermination. Then, I use inferences that Papineau himself accepts to argue that the disjunctive consequent of principle (G) is not met in the instance of overdetermination I have described.

Here is a description of a clear instance of overdetermination. Suppose a widely held variety of dualism is true: all mental states are irreducibly immaterial, but the mental supervenes on the material with nomic necessity. Add to this view that at least some mental states have *only one* subvenient base. That is, for at least one mental state M, there is only one material state that synchronically nomically necessitates M. Let's say that pain is one of these mental states, and let's say that C-fibers firing is pain's only subvenient base. Given the laws of nature, the only way to produce pain is for C-fibers to fire: no

C-fibers firing, no pain. It will follow on this view, then, that C-fibers firing is nomically sufficient *and necessary* for pain.

Now suppose that every physical event has a prior sufficient physical cause. Suppose also that conscious mental occurrences have physical effects. Those physical effects of conscious mental occurrences will therefore *clearly* be overdetermined: for each, there will be a prior sufficient physical cause and also a distinct prior sufficient mental cause. Consider, for instance, the pain that causes me to wince. On this dualist view, my wincing has two distinct sufficient causes: the pain and also the C-fibers firing. If this isn't overdetermination, nothing is.

So much for the clear case of overdetermination. Now turn your attention to the disjunctive consequent of (G). Using only inferences that Papineau himself accepts, I will now argue that even in this clear case of overdetermination the relevant instance of the consequent of (G) is false. Principle (G) will therefore be refuted by counterexample.

Focus on the wincing example I just mentioned. On this dualist view, C-fibers firing is nomically necessary for pain: it is pain's only subvenient base. On this view, then, it will be true that

> (A*) If C-fibers firing had not occurred, the pain would not have occurred.

Given the truth of this dualist view, worlds in which there are no C-fibers firing and yet there is pain are worlds with different laws of nature from ours. Surely, therefore, there are worlds that are 'closer' or 'more similar' to ours in the relevant senses in which there are no C-fibers firing and also no pain occurring, worlds that do not involve miracles. And therefore (A) is true.

Since he endorsed the inference from (A) to (B) above, Papineau should accept that (A*) entails

> (B*) It is false that I still would have winced even if the C-fibers firing had not occurred.

Similarly, since C-fibers firing is nomically sufficient for pain, it will be true on this dualist view that

> (C*) If the pain had not occurred, C-fibers firing would also not have occurred.

Given the truth of this dualist view, worlds in which there is no pain and yet there are C-fibers firing are worlds with different laws of nature from ours. Surely, therefore, there are worlds that are 'closer' or 'more similar' to ours in the relevant senses in which there is no pain and also no C-fibers firing, worlds that do not involve miracles. Therefore (C*) is true.

Since he endorsed the inference from (C) to (D) above, Papineau should accept that it follows from (C*) that

(D*) It is false that I still would have winced even if the pain had not occurred.

Putting (B*) and (D*) together, we get

(E*) It is false that (EITHER *I still would have winced even if the pain had been absent* OR *I still would have winced even if the C-fibers firing had not occurred*).

But this is the negation of the consequent of the relevant instance of general principle (G). And that is what we were trying to prove. Principle (G) is therefore refuted by counterexample.

We have described a view on which my wincing clearly involves overdetermination, and yet the wincing does not meet the necessary condition that Papineau proposes for overdetermination. Therefore the condition that Papineau proposes as necessary for overdetermination is not necessary. But this condition on overdetermination was the only reason that Papineau proposed in favor of thinking that role functionalism would not involve overdetermination using his generous sense of 'cause', and so would be consistent with his premises (1*)-(3). Therefore, Papineau has given us no good reason to believe that role functionalism is consistent with premises (1*)-(3).

But at the same time we have a good reason to think that role functionalism is *not* consistent with premises (1*)-(3): given the

generous sense of 'cause' involved in (1*), it comes out on role functionalism that the effects of conscious events will always have *two distinct sufficient causes*. Plausibly, if an effect E has two distinct sufficient causes C1 and C2, then E is overdetermined. Therefore, it will come out on role functionalism together with the generous sense of 'cause' involved in (1*) that the effects of conscious causes are always overdetermined by distinct causes. And that is inconsistent with premise (3). Therefore, role functionalism is inconsistent with premises (1*)-(3). This, recall, is the second horn of the dilemma for Papineau.

I have already shown how role functionalism is inconsistent with (1), a natural reading of the Causal Argument's first premise. We have now seen how role functionalism is inconsistent with (1*)-(3), where (1*) is a more generous reading of the Causal Argument's first premise. Since no other readings of that first premise are on the table, I conclude that role functionalism is inconsistent with Papineau's version of the Causal Argument for materialism.

This is a significant result, for many materialists believe that, *qua* materialists, their only live options are brain state identity theory, realizer functionalism, and role functionalism. These materialists are pushed toward role functionalism by considerations of multiple realizability. If other species with radically different neural structures

can nevertheless share our exact mental state types, then brain state identity theory and realizer functionalism are both false. And if those two theories are false, they reason, then role functionalism must be true. And if role functionalism is false, they contrapose, then multiple realizability is impossible.

Yet we have just learned that if Papineau's version of the Causal Argument is sound, then role functionalism is false. And therefore a materialist ought to conclude that if Papineau's Causal Argument is sound, multiple realizability is impossible. The apparent possibility of multiple realizability, therefore, provides the materialist with a rebutting defeater for Papineau's version of the Causal Argument. If a materialist is strongly inclined, as I am, to believe that other species with radically different neural structures can nevertheless share our exact mental state types, the materialist should reject Papineau's version of the Causal Argument for materialism.

In conclusion, we learned in the first section that Papineau's version of the Causal Argument is valid only if the physical effects of conscious causes are *never* overdetermined by distinct causes. If you think it possible that the physical effects of conscious causes are *at least sometimes* overdetermined, you should reject Papineau's argument as unsound.

My second objection showed that, from the perspective of many materialists, Papineau's version of the Causal Argument is sound only if other species with radically different neural structures *cannot* share our exact mental state types. If you are a materialist who finds plausible the multiple realizability of mental states, you should reject Papineau's version of the Causal Argument as unsound.

We have here, then, two new reasons to reject this Causal Argument as unsound. Even if we had not found these flaws in the argument, it is plausible that the strength of our dualist intuitions would still have justified us in rejecting the conjunction of Papineau's premises. But given the flaws we have discovered, we are justified to an even greater degree in rejecting the conjunction of Papineau's premises.

And what goes for Papineau's argument for non-dualism also goes for every other argument for non-dualism and objection to dualism. Therefore, we have here a general strategy for responding to proposed rebutting defeaters to dualism. Even if one cannot find any flaws in the proposed rebutting defeaters for dualism that I sketched above, so long as one finds Possibility and Non-Identity more plausible than the conjunction of the premises and inferences of each of the proposed rebutting defeaters, none of them will succeed. However, in many cases one might find a premise or inference that is

particularly dubious. Take Jaegwon Kim's Pairing Problem. First, the conclusion is consistent with a dualism about property types; it is only a problem for substance dualism. Second, the argument is a problem only for those varieties of substance dualism that accept that souls are not spatially located. The Pairing Problem should not trouble a dualist who is willing to jettison that assumption. Take also the inductive argument for non-dualism: it will rely on something like the assumption that there are no relevant differences between phenomena that are uncontroversially physical and mental phenomena. Anyone remotely attracted to Possibility and Non-Identity will have reason to reject that assumption. A similar response can be given to Stoljar's methodological argument for non-dualism. Now take the simplicity argument for non-dualism. It will rely on something like the assumption that non-dualism and dualism explain the data equally well, and thus are candidates for Occam's Razor. Again, anyone remotely moved by Possibility and Non-Identity will think that dualism has a serious advantage over non-dualism, namely that dualism does justice to these intuitions while non-dualism denies them. Zimmerman's objection to dualism relies on a small army of assumptions, many of which are less plausible than Possibility and Non-Identity, and the conjunction of which certainly is. These include the assumption that substance dualism is false. It is an argument meant to move the mere property dualist

towards substance dualism. A substance dualist will, therefore, be unmoved by this argument. Similarly with Adam's objection: it is meant to be an objection to atheistic dualism. A theist will not be moved by this argument. And a firmly atheistic dualist will likely find Possibility and Non-Identity more plausible than the conjunction of Adam's myriad other assumptions.

These are brief remarks on each argument, but for a natural reason. The general strategy requires only that we evaluate the conjunction of premises for any argument for non-dualism or objection to dualism. If that conjunction is less plausible than Possibility or Non-Identity, then the proposed rebutting defeater is defused. We need not locate any particular error; we need not know exactly where the argument goes wrong. It is sufficient to note that there must be an error somewhere, given the high probability of Possibility and Non-Identity vis-à-vis the conjunction of the rebutting defeater's premises. Locating especially dubious premises is an unnecessary bonus.

J.J.C Smart (1959, 143), an ardent non-dualist, agrees with the general methodology here. Speaking of the entities and relations posited by dualism, he says "If any philosophical arguments seemed to compel us to believe in such things, I would suspect a catch in the argument." Though our respective prior probabilities in dualism and

materialism differ significantly, we agree that, when confronted with an argument for a seriously improbable conclusion, the right attitude is to suspect a catch in the argument.

So it goes with powerful but abstruse arguments in which the conjunction of the premises and inferences deserves less credence than the negation of the conclusion. But could a philosophical argument ever be so powerful as to provide more than a merely diminishing rebutting defeater for a proposition that intellectually seems obviously true? To answer that, suppose that someone–perhaps inspired by the Pythagoreans–presents you with a subtle but extremely powerful argument for the conclusion that numbers are the ultimate reality, and that therefore

(5) The Prime Minister is a prime number.

Suppose that, try as you might, you can't find any particular fault with the argument. We have already established that if the conjunction of premises and inferences in the argument deserves less credence than the denial of the conclusion, then this argument does not give you an overriding or neutralizing rebutting defeater for your intuitive belief that

(6) It is *not* the case that the Prime Minister is a prime number.

Perhaps this Pythagorean argument would give you pause for thought, and perhaps you ought to continue to grapple with the

argument. Perhaps it would even give you a diminishing rebutting defeater. Yet in order to take this Pythagorean argument as evidence for (5), one must take the argument to have true premises and a valid inference, i.e. one must believe

(7) This Pythagorean argument is sound.

If one takes (6) to have more going for it than (7), then she does not have an overriding or neutralizing rebutting defeater for her belief that (6).

Could the evidence in favor of (5) and evidence in favor of (6) ever be on a par, such that the right attitude is to suspend judgment? Could there be circumstances in which this Pythagorean argument is a *neutralizing* rebutting defeater for (6)? Suppose that each premise of the Pythagorean argument is as intuitively obvious as (6), and that the inference from the premises of the Pythagorean argument to (5) is as intuitively obvious as (6). (Let's grant that we can't have Cartesian certainty about the premises of the Pythagorean argument, or the validity of the inference, or even (6), so let's say we judge their individual probabilities to be just less than 1.) It would be a remarkable achievement for any interesting philosophical argument to have premises and inferences each of which is as subjectively probable as (6). Let's say that under these conditions, this Pythagorean argument is a *knock-down* argument.

While the premises and the inference from premises to conclusion may individually be as intuitively obvious as (6), (7) itself will not be as credible as (6). After all, in order to take (7) as evidence for (5) in a way that could actually give (6) a run for its money, one must believe a conjunctive proposition, namely that each of the premises of the Pythagorean argument is true, and the inference from the premises to (5) is valid. And one must believe these conjuncts on the basis of intuition.[39]

But this conjunctive claim is logically stronger than each conjunct individually. So the credence that it deserves is a product of the probabilities of the individual conjuncts. And this product must be lower than the credence deserved by each conjunct individually.[40] And we've already supposed that each conjunct deserves as much credence as (6). Therefore, even if the Pythagorean argument were a knock-down argument, the evidence supplied by the Pythagorean argument in favor of (5) (taken as such) would still be inferior to our evidence in favor of (6), and therefore (6) would still not deserve as much credence as (6). Therefore, the Pythagorean argument could supply neither an overriding nor a neutralizing defeater for (6), even if the argument were a knock-down argument.

---

[39] As opposed to, say, believing (7) on the basis of testimony, in which case we certainly wouldn't have a neutralizing rebutting defeater.

[40] At least if you judge the conjuncts to be anything less than certain, i.e. if you judge their probability to be anything less than 1.0.

Let's now return to Papineau's Causal Argument. Papineau's argument is not a knock-down argument by any stretch of the imagination, but the last paragraph shows that *even if* all the premises in Papineau's argument and the inference from the premises to the conclusion were as intuitively obvious as Possibility and Non-Identity, still we would only get at most a diminishing rebutting defeater. The same line of reasoning would apply to any argument for any of the reductive type-identity theories. Importantly, the same line of reasoning would apply to any objection to dualism, i.e. any argument for the conclusion that dualism is false using commitments of dualism as premises.

Conclusion

In conclusion, then, someone to whom Possibility and Non-Identity seem obviously true can rest her dualism on firm ground. For such a person, even if there are powerful philosophical arguments for the conclusion that one of reductive type-identity views is true, or powerful philosophical objections to the truth of dualism, the fact that Possibility and Non-Identity intellectually seem so obvious, combined with the fact that this clear, intuitive seeming has not been undercut, make the standards for success for an overriding or even neutralizing rebutting defeater extremely high.

Even if the argument or objection were knock-down – i.e. the premises and inference from premises to conclusion are each as intuitively obvious as Possibility and Non-Identity – it would only give the dualist at most a diminishing rebutting defeater, and (supposing the initial subjective probability of dualism were sufficiently high) the right attitude to take would still be dualism. The only threat to the dualist is the dim prospect of an argument for non-dualism or an objection to dualism with a conjunction of premises and inferences that is more credible than the disjunction of Possibility or Non-Identity. At least in my case, none of the arguments we've surveyed meets this high standard, and none seems forthcoming.

Perhaps if we had any clear intuitions that a type-identity theory is true, then we would have a neutralizing rebutting defeater and the right position would be suspension of judgment. Or perhaps if we had an argument for a type-identity theory or an objection to dualism in which at least one premise (or the inference from the premises to the conclusion) were *more* intuitively obvious than Possibility or Non-Identity, then perhaps we would have a neutralizing or even overriding rebutting defeater for our dualist intuitions (depending on just how much more obvious the premise or inference were, and how many premises there were).

But we clearly don't have any intuitions that some reductive type-identity theory is true – none is just obvious. Philosophers tend to accept reductive type-identity theories on the basis of some philosophical argument.[41,42] But given the intuitive truth of Possibility and Non-Identity, it may well be that no such argument can justify belief in any reductive type-identity theory; no such argument that I have seen (or even heard the faintest rumor of) can even justify suspension of belief. The right attitude is therefore belief in dualism, even if we had access to a knock-down argument for a type-identity theory or a knock-down objection to dualism.[43]

---

41 Papineau agrees that it takes more than intuition to justify materialism (2002, 36-38). What's needed, he says, is an argument in favor of materialism.

42 Or perhaps Tyler Burge (1993) is correct when he says that the naturalistic picture of the world is more like a political or religious ideology than like a position well supported by the evidence, and that materialism is an article of faith based on the worship of science.

43 Perhaps no single proposed rebutting defeater can override our dualist intuitions. But what about the disjunction of all the proposals? Or what about the proposals considered as a series, each one chipping away at our confidence in dualism? Here things get complicated. It may well be that Possibility and Non-Identity strike you as maximally obvious—like the claim that pain isn't euphoria, or pain isn't the property of being a misaligned floorboard. If so, there will be no "chipping away" at your dualism. But suppose you take Possibility and Non-Identity as less than absolutely certain. It may well be that a sufficiently robust series of proposed rebutting defeaters could, in the end, override your dualism. I suppose this is what has in fact happened with those many non-dualists who admit to having dualist intuitions. Each of us must do the math in his or her own case, assigning probabilities to Possibility and Non-Identity, and then to the premises and inferences of each proposed rebutting defeater. In my case, dualism survives the onslaught. Perhaps the same goes for you.

CHAPTER FOUR

An Argument for Dualism from the Nature of Introspection

In the previous three chapters, we have developed a methodological argument for dualism. In this chapter, I will set that argument aside and present a novel argument for the view that I am not a complex thing, and so not a complex physical object like a brain or a body.

In 1962, Kurt Baier argued that materialism about the mind entails introspective fallibility, e.g. that you might be wrong that you are in pain at a time when it seems to you that you are in pain. Since very many philosophers at the time accepted introspective *in*fallibility, this was a significant result. In response, materialists like David Armstrong (1963) squared their shoulders, accepted the implication, and haven't looked back. These days, it is hard to find a philosopher who accepts that introspection is infallible.

Even so, most contemporary philosophers believe that introspection is not *hyper*fallible. Introspection might get it somewhat wrong, they think, but introspection can't get it radically wrong. It couldn't be, for example, that I am actually experiencing fierce pain

right now though it introspectively seems to me that I am not. In this paper, I will extend Baier's argument to target this popular view. I will argue, roughly, that each of the standard accounts of introspection on which it is *mechanistic*—that is, a causal series of events extended in time—entails introspective hyperfallibility. If any one of the standard accounts of introspection is right, then we *never* have introspective certainty—every last one of our introspective beliefs is defeasible. This implication, I take it, is far less palatable than the one Baier pointed out.

Furthermore, I will argue that any version of what I call "the Complex View"—including the standard materialist view on which people are complex material objects like bodies or brains—entails that introspection might be mechanistic. What follows, I will argue, is that the Complex View forces open the possibility that none of our introspective beliefs is certain for us. Philosophers who believe that is not an open possibility will consider this to be a powerful argument against the Complex View—they will think certainty teaches us that the standard materialist view of human persons is false, and that human persons are simple.

1. The Main Argument

I am some thing. Many people think that I am a complex thing—a thing with parts—and that my mental life is (or is a result of) the interaction of some of these parts. Which complex thing am I? Perhaps I am a body, or perhaps some part of a body such as a brain, or perhaps some special part of a brain. Other people think that I am not a complex thing at all. Rather, these people say that I am a simple thing—a thing with no parts—and my mental life is a basic activity of this simple thing, not a result of the interaction of any parts.[44]

In this paper, I will develop and defend a novel argument that may be used to support the Simple View:

> (1) Supposing that the Complex View is true, I cannot be certain[45] that introspection[46] is not a causal series of events extended in time.

---

[44] See, for example, Roderick Chisholm (1991) and, more recently, David Barnett (2010). Earlier in his career, Chisholm (1978) took seriously the possibility that we are material simples, though he stopped short of endorsing the view.

[45]. A proposition *p* is certain for a subject S just in case S is entitled to believe *p*, come what may. That is, S's grounds for *p* make it such that no additional evidence should lower her credence in *p*. By "*p* is certain for S" I do *not* mean "it is psychologically impossible for S to doubt that *p*." Rather, I mean the normative notion "S cannot rationally doubt *p*." I mean what many have called "absolute" or "Cartesian" certainty. When a subject has this kind of certainty, her belief is often said to be "Demon-proof" after Descartes' *deus deceptor*.

[46]. There are many ways a subject might come to have beliefs about the phenomenal character of her own experiences. Introspection is that way to which the subject has privileged access (normally, at least).

(2) If introspection is a causal series of events extended in time, then I could gain good evidence that I am feeling fierce pain right now.[47]

(3) For any propositions *p* and *q*, if (*i*) I cannot be certain that *p* is false, and (*ii*) *p* entails *q*, then I cannot be certain that *q* is false.

(4) So, supposing that the Complex View is true, I cannot be certain that I could not gain good evidence that I am feeling fierce pain right now.

If you think it is more likely that *the Complex View is false* than that *you cannot be certain that you couldn't gain good evidence that you're in fierce pain*, then (1)-(4) constitute an argument for the conclusion that the Complex View is false. Let me now explain and motivate the premises and then respond to three objections.

2. Introspection as a Causal Series of Events Extended in Time

Visual perception is a process, a series of events extended in time whereby one comes to have beliefs about how the world is. We commonly take this to be a *causal* series of events extended in time: each member of this series of events is merely nomically sufficient for the next, and each member of this series could

[47]. That is, I could gain evidence that would make it reasonable for me to believe that I am experiencing fierce pain, even though I continue to token just this type of experience, which introspectively seems to me not to involve any pain at all.

occupy a point or duration of time distinct from that of any other member in this series. For example, we think some story like this is now true of you: There is a surface with markings before you. This surface and these markings cause light to be reflected in a certain way into your eyes. This reflected light causes certain events on your retinas. These retinal events cause certain events in your optic nerve. These optic nerve events cause certain events in your visual cortex. Then you enjoy a visual experience, which represents (among other things) that there is a white surface with black markings before you.

Like visual perception, introspection is a process, a series of events extended in time. Unlike visual perception however, introspection is a process by which one becomes aware not of the external world, but rather of the phenomenal character of her own experiences. For example, we think some story like this is now true of you: You have a visual experience, which represents that there is a white surface with black markings before you. Then you attend to some of the phenomenal character of your experience – the whiteness, say. And then, somehow, you end up with the belief (or awareness, or perception) that you are having a visual experience as of white. Many people think that introspection, like visual perception, is a *causal* series of events, involving some sort of *mechanism*. For example, Alex Byrne (2005) writes: "[U]nless it's

magic, I must have some sort of mechanism (perhaps more than one) for detecting my own mental states—something rather like my visual, auditory, and gustatory systems, although directed to my mental life." I will first discuss some mechanistic views of introspection before arguing for the premises of the Main Argument.

3. Some Mechanistic Views of Introspection

My introspective awareness that I am having a visual experience as of white stands in some relation to that visual experience itself. There are many views of introspection on which some causal mechanism takes the first-order state as input and delivers the introspective state as output. On these views, some sort of temporally extended causal chain leads from the first-order state to the introspective state.

David Armstrong's and William Lycan's Inner Sense Model of introspection is one of these views. According to Lycan (2003) "...introspection is the operation of an internal attention mechanism that monitors experiences and produces second-order representations of their properties..." These second-order representations are importantly similar to ordinary *perceptions*, and thus this view has become known as the Higher-Order

Perception (HOP) view of introspection. Lycan says introspection makes us aware of our experiences and their properties, as perception makes us aware of external objects (like bottle rockets) and their properties. That is, introspection is a mechanism that delivers second-order *perceptions* that there is an experience that there is such and such, just as visual perception is a mechanism that delivers first-order *perceptions* that there is such and such. For my purposes, it is important to note only that (*i*) if the Complex View is true, I cannot be certain that introspection does not work this way, and (*ii*) on this view introspection is a temporally extended causal chain—mediated by this internal attention mechanism—leading from experiences to distinct second-order representations of their properties.

Consider now a Higher-Order Thought (HOT) view of introspective awareness advocated by Sydney Shoemaker (1994), David Rosenthal (2004), Shaun Nichols and Stephen Stich (2003), among many others. On this view, introspection is the process by which we come to have non-perceptual second-order self-attributions about our first-order mental states. According to Shoemaker, these second-order states are beliefs. And the brain state associated with the first-order mental state *causes* the brain state associated with the belief about it. According to Rosenthal, a mental state is conscious only if it is accompanied by a distinct,

occurrent HOT. Right now, I am conscious that there is a white surface before me; the visual representation is accompanied by that HOT. In introspection we become conscious of our consciousness; that HOT itself comes to have—via a causal process extended in time—an accompanying HOT. I become conscious that I am conscious that there is a white surface before me.

A variant HOT theory was put to me by Michael Tye, though I do not know how seriously he takes it. On what we may call a "Read-Write Model," there is a consciousness-compartment (C-box) in the mind, in addition to a belief-compartment (B-box). When one's experience represents that *p*, the sentence "*p*" (in the language of thought) is inscribed in the C-box. Introspection is a mechanism and one of its jobs is to read sentences inscribed in the C-box and write corresponding sentences in the belief-box, e.g. "I am aware that *p*." For my purposes, it is important to note only that (*i*) if the Complex View is true, I cannot be certain that introspection does not work this way, and (*ii*) on all these variations of HOT theory, introspection is a temporally extended causal chain leading from some type of first-order state to a distinct second-order representation of it.

Finally, let us consider Same-Order Monitoring Theory (SOMT) advocated by Uriah Kriegel,[48] and by some accounts Franz Brentano. According to Kriegel (2007, 370), a visual experience as of green (call that mental state "M") is a "complex" of a visual representation of green (call that "M1") bundled with the awareness of M1, i.e. an appropriate representation of M1 (call this state "M2"). In virtue of being represented by M2, M1 is conscious, and M is a visual *experience* as of green rather than a mere representation. Kriegel seems to agree that the first-order state does not represent itself *as* being represented. Rather, that is what introspection does.

There are a few plausible ways introspection might work on Kriegel's view. It may be that a higher-order representation either of M2 or of M is not part of the complex M, but is brought about by some causal mechanism. Alternatively, it may be that the complex target state M *comes to have as a constituent* a representation of M2, via some causal mechanism (cf. Rosenthal 2004, 33). For my purposes, it is important to note only that (*i*) if the Complex View is true, I cannot be certain that introspection does not work this way, and (*ii*) on any of these plausible SOMT views of introspection, introspection is a temporally extended

[48]. What I say here is also applicable mutatis mutandis to Intrinsic Higher-Order Thought Theory advocated by Genarro (1996) and Natsoulas (1996).

causal chain leading from some type of first-order state to a distinct representation of it.

4. Support for Premise (1) in the Main Argument

Recall the first premise of the Main Argument:

> (1) Supposing that the Complex View is true, I cannot be certain that introspection is not a causal series of events extended in time.

Let me now support this premise. I take it that none of the views discussed in the previous section is obviously false, at least on the assumption that I am a brain or some other complex object. After all, if I suppose that I am a brain and that my mental states supervene on the physical states of that brain, then it may be that the physical states on which the mental states that constitute introspection supervene are a temporally extended causal chain. If so, it may be that the supervening mental states that constitute introspection are a temporally extended causal chain. The theories discussed in the previous section are reasonable explanations of how introspection might work, given these assumptions.

To put it somewhat more vividly, suppose you are a brain. Given that assumption, there is evidence I could give you to make it reasonable for you to believe that introspection works as, for

example, the Read-Write Model suggests. Say I pop open your skull and show you the goings-on therein: if you just are that brain, and your mental life is intimately related to various events in that brain, couldn't it be that there is, say, a read-write introspective mechanism in there? On the assumption that you are a brain, you cannot rule out this theory from the armchair—this theory is clearly broadly logically possible. And the same goes for the other theories of introspection as well, each of which suggests that introspection is a temporally extended causal chain. On the assumption that the Complex View is true, none of these theories is a priori knowably false. You may believe that one or more are false, but this belief can't be *absolutely* certain for you. Since this is all premise (1) in the Main Argument claims, we should accept it.

5. Support for Premise (2) in the Main Argument

> Recall the second premise of the Main Argument:
>
> (2) If introspection is a causal series of events extended in time, then I could gain good evidence that I am feeling fierce pain right now.

Let me now support this premise. Consider first visual perception. Because visual perception involves a causal series of events extended in time—because, that is, it is mechanistic—it is subject

to radical correction and *none* of its deliverances is certain for you. After all, it is metaphysically possible for any causal series of events to go (very) awry, and to lead to a (very) statistically abnormal, improper, or inapt result.[49] And in any case of visual perception, I could present you with evidence that would make it reasonable to believe that the causal chain has in fact gone (very) awry, and that your visual experience (really badly) misrepresents the way the world is.

For example, though your current visual experience represents that there is a white surface with black markings before you, there is evidence I could give you that would make it reasonable for you to believe that your mechanism of visual perception has malfunctioned, and that actually there is only a red surface with green markings before you (so the causal chain has gone awry), or that actually there is no surface and no markings at all (so the causal chain has gone *very* awry). The well-rehearsed stories involve the usual suspects: malevolent neurosurgeons from Alpha Centauri, an Evil Demon, hallucinogenic drugs, etc. In general, since visual perception is mechanistic, for any possible visual experience E you may have, though E represents that the

[49]. Michael Tooley (2008, 97), I think, would agree. He says: "assuming that at least some of the basic causal laws of our world are probabilistic, any physical structure is capable of not functioning properly, and so any capacities based on a physical structure could always fail."

world is a certain way, it could be that the world is different from how E represents it to be, even radically different. No matter how things visually seem, we recognize the possibility that things are not as they seem. Things may even be *very* different from how they seem to us, if someone is tampering with our visual mechanisms in the right way.

And why should the same inference not hold in the case of introspection, if it too is mechanistic?[50] If introspection is a causal series of events extended in time, and any causal series of events could go (very) awry, introspection is also subject to radical correction and none of its deliverances is certain for you. On any occasion of operation, the physically-realized introspective mechanism could malfunction, and could deliver (very) false self-ascriptions, second-order beliefs, second-order perceptions, or whatever output your favored theory suggests. If introspection is only in the business of delivering the awareness or thought or perception or belief that *I am experiencing that p*, then if I come to believe that my introspective mechanism is malfunctioning, or that the causal process has gone awry, then the deliverances of

---

[50]. D.M. Armstrong (1963) seems to think it does: "I shall defend the thesis…that mental states are…states of the brain. Now if I accept the existence of introspection, as I also do, then I must conceive of both introspection and the objects of introspection as states of the brain. Introspection must be a self-scanning process in the brain. That it is logically possible that such a self-scanning process will yield wrong results is at once clear…"

that mechanism are subject to radical correction. Just as I accept the possibility that things are very different from how they visually seem (since visual perception is mechanistic), I ought to accept the possibility that things are very different from how they introspectively seem, if introspection is mechanistic.

Consider the Read-Write Model of introspection discussed above. Assuming it is operating according to a good design plan, if the mechanism is functioning properly and reads the sentence "*p*" in the C-box (i.e. the Consciousness-Box), it writes "I am aware that *p*" in the B-box (i.e. the Belief-Box). But it is in principle possible to manipulate the mechanism such that it is no longer functioning properly, such that it for example reads "*p*" in the C-box and writes "I am aware that **not**-*p*" in the B-box. Assuming that my brain realizes this mechanism, a sufficiently clever neurosurgeon could in principle manipulate my introspective mechanism in this way.

Therefore, according to the Read-Write Model, on any occasion in which my introspective mechanism has inscribed "I am aware that **not**-*p*" in my B-box, I could gain evidence which would make it reasonable for me to believe that I am actually aware that *p*. It might go like this, on the assumption that the Complex View is true: first, I gain evidence that makes it

reasonable to believe that I am the victim of a fiendish neurosurgeon. Then, I gain evidence that makes it reasonable to believe that I have an introspective read-write mechanism, and that this neurosurgeon is causing it to malfunction in so that, though *there is fierce pain* is written in my C-box, only *I am aware that it's not the case that there is fierce pain* is written in my B-box.[51] In such an instance, it would be reasonable for me to believe that my introspective beliefs are radically false. Though it would surely introspectively seem that I am not in fierce pain, in this case I would have good reason to believe that things are not as they introspectively seem.

And so it follows that, on the Read-Write Model, I could gain evidence that would make it reasonable for me to believe that I am feeling fierce pain right now. Similar considerations apply to the other versions of HOT including Shoemaker's model,[52] to the

---

51. If one is concerned with immaterialist versions of the Complex View, the evidence gained here could be about an Evil Demon rather than a neurosurgeon.

52. On Shoemaker's view, it may be that the experience of pain is such that, *in certain circumstances*, it necessarily causes the second-order self-attribution *I am aware that there's pain*, and it may be such that this second-order self-attribution is such that, *in the absence of malfunction*, it is caused by the first-order state. But a sufficiently clever neurosurgeon could manipulate one's brain to produce malfunction, to produce a circumstance that is not one of those in which the experience of pain necessarily causes the second-order self-attribution. Thus the neurosurgeon could manipulate my brain such that, though my experience represents that *p*, I form via introspection the belief that *I am <u>not</u> aware that p*.

Inner Sense Model,[53] and to SOMT.[54] In fact the point generalizes to any view of introspection according to which it is a causal series of events extended in time. And therefore we should accept premise (2) of the Main Argument. Having now supported that premise, let me move on to premise (3).

6. Support for Premise (3) in the Main Argument

Recall the third premise of the Main Argument:

> (3) For any propositions *p* and *q*, if (*i*) I cannot be certain that *p* is false, and (*ii*) *p* entails *q*, then I cannot be certain that *q* is false.

First, a preliminary note about "entails" as it appears in (*ii*). For the move from (1) and (2) to (4) to be valid, the "entails" in clause (*ii*) of premise (3) must refer to whatever sort of entailment

---

[53]. When the Inner Sense Model's internal scanner is functioning properly (assuming a good design plan), if my first-order experiential state has the quale P, then the scanner will produce a second-order representation which, while not *itself* having quale P, represents that the first-order state has quale P. However, it is in principle possible to manipulate the mechanism such that, even though my first-order experiential state has the quale P, the second-order representation produced by the scanner represents that the first-order state does not have quale P.

[54]. However introspection works on Kriegel's view, if the introspective mechanism is functioning properly (assuming to a good design plan), if I am visually aware that *p*, my introspective mechanism will produce a representation in virtue of which I am introspectively aware that I am visually aware that *p*. However, this mechanism may be manipulated such that, even though I am visually aware that *p*, it produces a representation in virtue of which I am introspectively aware that I am not visually aware that *p*.

relation is expressed as holding between the antecedent and consequent of premise (2). I take it that there is no algorithmic way of settling the claim made by (2), as there is, by contrast, with claims of first-order entailment. In this way, the consequence relation claimed by (2) is akin to the relation claimed by the proposition that *for any x, if x is a prime minister, x is not a prime number.* I take it that we have epistemic faculties that at least *can* deliver certainty regarding matters such as these, matters which we have no algorithmic method of settling.[55]

If so, then (3) can be proven indirectly: assuming that (3) is false results in a contradiction. To see this, suppose first that you cannot be certain that some proposition *p* is false, i.e. that your epistemic faculties cannot deliver certainty that *p* is false. Suppose further that *p* entails some other proposition *q*. (You may or may not believe that *p* entails *q*.) Now suppose that, contrary to (3), you *can* be certain that *q* is false.

I take it to follow obviously that in such a case you at least *can* be certain that *p* is false. All it would take is for your epistemic faculties to deliver certainty that *p* entails *q*, and certainty of modus tollens. You may as a matter of fact not realize that *p*

[55] Just to be crystal clear, "entails" as it appears in (3) is not equivalent to mere material implication. As we know, material implication is a sad model of genuine entailment.

entails *q*, and you may not believe that *p* is false. Nevertheless it is true that you *can* be certain that *p* is false. But then we stumble onto a contradiction. Attempting to construct an instance in which the antecedent of this conditional is true while the consequent is false results in absurdity. And so we should accept that (3) is true.

Consider also the following proposition, which is logically equivalent to (3):[56]

> (3*) For any propositions *p* and *q*, if (*i*) I can be certain that *p* is true, and (*ii*) *p* entails *q*, then I can be certain that *q* is true.

Think about an instance in which the antecedent is true: for some *p* and *q*, *p* entails *q* and your epistemic faculties at least *can* deliver certainty that *p* is true. In this case, you cannot be certain that *q* is true only if your epistemic faculties cannot even in principle deliver certainty that *p* entails *q*, or certainty of modus ponens. Yet surely you at least *can* be certain of those things. So we should accept (3*) and its equivalent: (3) itself.

7. What to Do with (4) in the Main Argument

Premise (4) follows from premises (1)-(3):

---

[56]. The contrapositive of (3) is this: For any *p* and *q*, if I can be certain that *q* is false, then either <it's false that *p* entails *q*> or <I can be certain that *p* is false>. Now let *p* represent *that q is false* and let *q* represent *that p is false*. (3*) is now obviously equivalent.

> (4) So, supposing that the Complex View is true, I cannot be certain that I could not gain good evidence that I am feeling fierce pain right now.

What (4) tells me, substantially, is that either the Complex View is false or I can't be certain that my belief that I am not experiencing fierce pain right now is indefeasible. I cannot rationally deny both of these; at least one is true. Which option I take should be determined by which I find more credible. If I find the antecedent of (4) more credible than the negation of the consequent, I should run a modus ponens. If on the other hand I find the negation of the consequent more credible than the antecedent, I should run a modus tollens. If I find the antecedent and the negation of the consequent equally credible, I ought to remain agnostic.

For what it's worth, I am strongly inclined to deny the consequent. There is very little I find more credible than that I can be certain that my belief that I am not experiencing fierce pain right now is indefeasible. Though I'm occasionally bothered by skeptical arguments aimed at every bit of my knowledge of the external world, I am never bothered by parallel skeptical arguments aimed at every bit of my knowledge of the inner world, so to speak. I am supremely confident that I could not receive any compelling evidence that, contrary to appearances, I really am experiencing fierce pain right now. My experience may

change—I may suddenly step in a rusty bear trap, for example—but given my current experience, surely nothing could defeat my belief that I am not experiencing fierce pain. I am far more confident of this than I am of the suggestion that, for example, I am a brain and my mental life is (or is a product of) the interaction of some of that brain's parts. And so I have a powerful argument against the Complex View, and in favor of the Simple View.

8. Objections

8.1 Tu Quoque

I use the Main Argument to support the Simple View. Some readers have suspected that I richly deserve a tu quoque response, since to them the inference in premise (1) seems equally valid in the case of the Simple View. That is, these objectors urge the plausibility of:

> (1*) Supposing that the *Simple* View is true, I cannot be certain that introspection is not a causal series of events extended in time.

And if (1*) is plausible, then of course the Main Argument could be turned against the Simple View. If (1) and (1*) are equally plausible, then whatever motivation the Main Argument

originally produced for the Simple View is neutralized by this revised argument.

My response is that (1*) is not plausible, or at least not as plausible at (1). The reason, ultimately, is that there is nothing about the Simple View that forces open the possibility of mechanistic introspection, while the same is not true of the Complex View. Mechanistic introspection may be ruled out from the armchair on the Simple View, but not on the Complex View.

Let me show you how. First, we will suppose that the Simple View is true. The Simple View does not have much content—it is just the denial of the Complex View. Now, our best bet to rule out the possibility of mechanistic introspection is, I believe, by running a modus tollens on something like premise (2) in the Main Argument. It would go like this: clearly, if introspection were mechanistic, then my belief that I'm not in fierce pain right now would be defeasible—all I would need is evidence that the mechanism malfunctioned. But since my belief that I am not in fierce pain right now is clearly indefeasible, it follows that introspection is not mechanistic.

Notice well that, as we close the door on the possibility of mechanistic introspection on the Simple View, it freely glides shut. Nothing in the nature of the Simple View forces open the

broadly logical possibility of mechanistic introspection. The Simple View is light on content, and so it presents no obstacle to a modus tollens on premise (2). And so, in this way, one in fact can be certain that introspection is not mechanistic, on the Simple View. Therefore, (1*) is implausible and—if the original premise (1) of the Main Argument is more plausible—the tu quoque objection fails.

Premise (1) in the Main Argument says that, if the Complex View is true, then we can't be certain that introspection is not mechanistic. At this point, you may be wondering whether premise (1) in the Main Argument really is plausible, and whether I didn't just sneak it by you back there in §4. After all, you might think, if it is so easy to be certain that introspection is not mechanistic on the Simple View, couldn't we likewise gain that certainty on the Complex View?

I think not, and here's why. Let's try to be certain that introspection is not mechanistic, on the Complex View. First, we will suppose that the Complex View is true. Now—in contrast to the Simple View—the Complex View comes with heavy baggage. It entails that each of us is a complex thing—a thing with parts—and each of our mental lives is (or is a result of) the causal interaction of some of these parts. Now, again, our best bet to rule

out the possibility of mechanistic introspection is, I believe, by running a modus tollens on something like premise (2) in the Main Argument. But notice here that, as we try to shut the door on the possibility of mechanistic introspection, we encounter some firm resistance: the nature of the Complex View gets in the way, insisting on the possibility of mechanistic introspection. For consider the set of possible worlds in which we are complexes like brains, and our mental lives supervene on the causal interaction of our parts. Isn't it obvious that, within this set of worlds, there are worlds in which introspection is a causal series of events extended in time? Consider also that, on the Complex View, I already know that other of my physically-realized belief-producing processes *are* mechanistic, e.g. perception. My introspective beliefs *may* result from a similar mechanism in a brain, if I just am a brain. We cannot shut the door on that broadly logical possibility from the armchair, given the hypothesis that I am a complex object like a brain.

In sum: aided by something like the Main Argument, one can come to see clearly that mechanistic introspection is impossible, in a way perfectly compatible with the Simple View. And so (1*) is not plausible. On the Complex View, however, mechanistic introspection clearly seems broadly logically possible, and so it cannot be ruled out from the armchair. And so (1) is

quite plausible. Therefore, since (1*) is less compelling than (1), the views are not on a par, contrary to what the tu quoque objection insists. Reflection on introspective certainty provides, therefore, a compelling argument against the Complex View, an argument that cannot return the favor to the Simple View.

One final thought. The main argument of this paper can be recast in terms of Bayesian confirmation. What the argument shows is that introspection cannot be mechanistic. We can compare the probability of this proposition conditional on the Complex View and conditional on the Simple View. How likely is it that introspection should not be mechanistic, assuming that the Complex View is true? Quite unlikely, I should think. On the other hand, how likely is it that introspection should not be mechanistic, assuming that the Simple View is true? Not nearly as unlikely, I should think. But then we have confirming evidence for the Simple View and disconfirming evidence for the Complex View.

8.2 Constitution and Incorrigibility

It has been suggested in the philosophical literature that the phenomenal character of a subject's experience is "taken up" into her corresponding introspective beliefs—that some of those

very abstract objects that constitute the representational content of her visual experience also (at least partially) constitute the representational content of her introspective belief about that experience. Some say phenomenal concepts are "quotational": they are said to "include" or "contain" the very phenomenal properties they refer to (see, for example, Chalmers 2003 and Block 2006).

Assuming sense can be given to its metaphors, such a theory would presumably secure an incorrigibility thesis along the lines of the one discussed in Jackson 1973: That *S believes at t that he is in pain at t* via introspection broadly logically guarantees that *S is in pain at t*. Similarly, that *the Statue of Liberty is on the pedestal at t* broadly logically entails that *matter is on the pedestal at t*. And, importantly, introspection could work this way and yet be a temporally extended causal chain of events.

And so an objector might urge that premise (2) is false, saying "Look, here's an account of introspection according to which it is a temporally extended causal chain, and yet according to which one may be certain about some introspective beliefs. Given the constitutive relation between the experience and the belief, the subject's belief has a very high epistemic status—the belief couldn't be false. And so this belief may rightly be said to be

certain for the subject. No evidence would make it reasonable for her to believe that she is in fierce pain."

The objection includes something like the following steps:

> ***Constitution***: Introspective beliefs are at least partly constituted by the states they are about.

Therefore,

> ***Incorrigibility***: That S introspectively believes she is in phenomenal state P at time t broadly logically guarantees that S is in P at t.

Therefore,

> ***Certainty***: We have an explanation of the certainty of at least some introspective beliefs.

And,

> ***Compatibility***: This explanation is compatible with mechanistic introspection.

Therefore,

> ***Counterexample***: Premise (2) in the Main Argument is false.

In what follows, I will give three critical responses to this objection.

My first response to this objection is that the inference from Certainty and Compatibility to Counterexample is invalid. Even if we grant that this constitution story is compatible with mechanistic introspection and that it can explain the certainty of introspective beliefs like *I am aware that there's fierce pain,* how does the story go with respect to what we may call "negative" introspective beliefs such as *I am aware that there's no fierce pain*? Right now, I have that belief and it is true. Yet it cannot be that some constituent of my experience is "taken up" into the introspective belief, since the introspective belief accurately represents that a certain phenomenal quality, namely fierce pain, is *absent* from the content of my experience.[57] And so the objector has not yet provided an account according to which introspection is mechanistic and yet I can be certain that I am aware that there's no fierce pain. And so premise (2) is unchallenged, since the proposed constitution story does not apply to negative introspective beliefs. Indeed, my Main Argument intentionally concerns a negative introspection belief in order to sidestep just such an objection.

My second response is that a theory that entails anything like Incorrigibility has highly implausible consequences. Such a

[57]. John Pollock (1986, 32-33) discusses this problem for constitution theories, i.e. theories which endorse what he calls the "Containment Thesis."

theory would entail, for example, that the following case is impossible:

> **Paint Store:** I am at a paint store, looking at samples. I hold a maroon color sample in the center of my visual field, which thereby tokens only maroon. Nevertheless, I misidentify the color and believe that I am visually experiencing scarlet in the center of my visual field.

Surely this story is coherent.[58] (My wife testifies that this is a common occurrence in my own life.) Yet, according to this constitution theory, that *I introspectively believe that I am experiencing scarlet in the center of my visual field* broadly logically guarantees that *I am experiencing scarlet in the center of my visual field*. So, if this constitution theory is right, Paint Store is incoherent. Yet since Paint Store is coherent, we should reject this constitution theory. And so this theory cannot offer us an objection to the Main Argument.[59] The objection stalls at the first

---

[58] Alan Goldman (2004, 282) agrees: "If one is inattentive, drugged, or otherwise imbalanced, he can misapply concepts to his own experiences and generate false beliefs about them." Goldman there also cites a case from John Pollock (2001, 43), saying: "...people typically and inattentively describe ways shadows appear on snow as gray, since they simply assume that shadows are gray, when in fact they appear blue."

[59]. Perhaps a version of the constitution theory survives, restricted to introspective beliefs like "I am aware of *this*," or the kind of beliefs recently discussed by Horgan and Kriegel (2007), or Chalmers' (2003) "direct phenomenal beliefs." But none of these theories furnishes an objection to the Main Argument, since none of them entails that my introspective belief that *I am not experiencing fierce pain right now* is incorrigible. Such theories explain, at most, the introspective certainty of some conceptually stripped-down, relatively content-less

steps: Incorrigibility above is false, and therefore so is Constitution, and so again the objection does not reach the conclusion that premise (2) in the Main Argument is false.

My third and final response is that nothing in the neighborhood of Incorrigibility is sufficient to explain the kind of epistemic certainty that attends some of our introspective beliefs. That is, for example, the inference from *if S believes she's in pain at t then she is in pain at t* to the conclusion that *S is certain that she's in pain at t* is invalid, and so the move from Incorrigibility to Certainty above fails. For consider that the number of hairs on your head is either even or odd. Suppose the number of hairs on your head is even here in the actual world, @. Given this supposition, it is true that if you believe the number of hairs on your head is even in @, then the number of hairs on your head is even in @. And that conditional is necessarily true, given that the proposition in question is a true world-indexed proposition. Across modal space, any creature with that belief believes truly; we have here necessary reliability and, indeed, incorrigibility.

And yet, despite the incorrigibility of this belief, it scarcely follows that this belief would be certain for you, were you to hold

---

introspective beliefs. And so they won't help explain the certainty of an introspective belief like *I am not in fierce pain right now*, which has substantially more conceptual content. It features the concept PAIN, for example, which none of the beliefs that concern Horgan and Kriegel or Chalmers do.

it. You would not be entitled to believe it, come what may. You lack any grounds at all for that belief, let alone the absolutely indefeasible grounds required for certainty. Evidently, the sort of incorrigibility thesis that the constitution story secures has little or nothing to do with justification and certainty. It follows, therefore, that even if an incorrigibility thesis in this neighborhood were true—and I argued above that it isn't—it would not by itself be enough to explain how I might be certain that I am not in fierce pain right now even if introspection is mechanistic. Since the move from Incorrigibility to Certainty above is invalid, the objection to premise (2) in the Main Argument again fails.

8.3 Constitution and Self-Intimation

In the previous section, we considered a theory of introspection on which introspective beliefs are partly constituted by the first-order states they are about. Such a theory would secure an incorrigibility thesis, as we said. But one might also wonder whether the constitution relation holds in the other direction. That is, one might wonder whether first-order phenomenal states like fierce pains are themselves partly composed of introspective awareness or belief. This view is not

unprecedented in the literature,[60] and it would secure an intuitively attractive self-intimation thesis: necessarily, if a subject is in fierce pain, then she's aware that she is.[61] And this view could be happily married with the suggestion from Lewis (1972, 258) that while the state S which normally plays the pain role *might* not be followed by the state S* that plays the awareness-of-pain role, under such conditions S would fail to satisfy our commonsense platitudes about pain and would therefore not be the referent of our theoretical term "pain."

It is not immediately obvious how such a self-intimation thesis might challenge the Main Argument. Exactly which premise(s) would it call into question? I believe it holds most promise of providing a counterexample to premise (2) in the Main Argument: an account of mechanistic introspection on which,

---

[60] Views like this are discussed in Weatherson 2004 (379) as well as Horgan and Kriegel 2007. Shoemaker (1990) thinks that at least some phenomenal states are "constitutively self-intimating," saying (2001) "it is of the essence of a state's having a certain phenomenal character that this issues in the subject's being introspectively aware of that character, or does so if the subject reflects." I'm concerned with this type of view in the text. Yet in other places, Shoemaker is careful to add the qualification that self-intimation doesn't occur with broadly logical necessity, but only under normal conditions, or absent malfunction. Such a qualified view would presumably be much less helpful to our objector here, since one's belief that one is properly functioning is far from certain. For that reason, I don't discuss this weaker, more qualified view here in the text.

[61] Or at least she would be, were she to introspect. I thank an anonymous referee for this journal for urging me to consider a constitution theory along these lines, and its implications for my Main Argument.

nevertheless, my belief that I'm not in fierce pain right now is indefeasible.

The objector may reason this way: "Listen, on this view I just sketched for you, it is obvious that, necessarily, if a subject is in fierce pain, she is aware that she is (or at least she would be, were she to introspect). Since by introspection you can be certain that you don't believe that you are in fierce pain, you can now run a modus tollens with that self-intimation thesis and conclude—with certainty—that you are not in fierce pain. This account of introspection secures the self-intimation thesis and thereby explains the certainty of your belief that you are not in fierce pain, all the while being compatible with mechanistic introspection. Therefore, premise (2) in the Main Argument is false."

This is a clever and interesting objection to the Main Argument. However, it passes the buck in a way that renders it, in the end, unsuccessful. Let's retrace the dialectic. Initially, I was wondering how I might be certain that I am not in fierce pain, if introspection is mechanistic. The objection now under consideration begins with this advice: "in order to be certain that you are not in fierce pain, start by being certain that you don't believe you are in fierce pain." But, of course, my initial concern will reemerge at this higher level—the problem has been merely

kicked upstairs, and the Main Argument can be redeployed against this new introspective belief. For how might I be certain that I don't believe that I am in fierce pain, if introspection is mechanistic? The same considerations about the fragility of causal process that first led me to worry about the epistemic status of my introspective belief that *I'm not in fierce pain* apply just as strongly to my introspective belief that *I don't believe that I'm in fierce pain*. Sure, it *seems* like I don't believe that I am in fierce pain, but if introspection is mechanistic there remains a possibility—remote, to be sure, but real nonetheless—that my introspective mechanism is malfunctioning, delivering the belief that I don't believe I am in fierce pain, when I really *do* believe that. Therefore, the objection has not yet explained how I may be certain that I am not in fierce pain right now, if introspection is mechanistic.

In other words, the objection promises me certainty that I am not in fierce pain, but only if I first gain certainty that I don't believe I am. Yet the objection does not explain how I might be certain that I don't believe I am in fierce pain. In this way, the objection passes the buck in an unsatisfactory way; it writes me a check that I cannot cash. And so, ultimately, the objection does not succeed in explaining how I might be certain that I am not in fierce pain even if introspection is mechanistic. And so, in the end,

it does not succeed in providing a counterexample to premise (2) in the Main Argument.

Secondly, this objection has it that my justification for believing that I am not in fierce pain right now is *inferential*, and in addition that this inference crucially depends on my knowledge of the self-intimation thesis. This strikes me as implausible on both counts: first, many people lack a belief in the self-intimation thesis, and yet are nevertheless certain—entitled to believe, come what may—that they're not in fierce pain right now. Children, for example. Second, even supposing that a person grasps and believes the self-intimation thesis, he may lack the cognitive resources to perform the somewhat complicated inference from that belief and his introspective belief that he doesn't believe he is in fierce pain to the conclusion that he is not in fierce pain. And yet, nevertheless, such a person may be certain that he is not in fierce pain; his grounds may make it such that no additional evidence should lower his credence. Children, again, serve as examples here. Since this objection has these two implausible consequences, we should again reject it as ultimately unsuccessful.

8.4 Nothing is Certain

I take the Main Argument to provide evidence against the Complex View. But that final conclusion is warranted only if it's true that some of our introspective beliefs are genuinely indefeasible. I assume that this is true, and that *I am not in excruciating pain right now* is such a belief. This assumption might be called into question, however. For example, Hartry Field (2000, 118) claims that "the credibility of any proposition could be diminished by evidence that well-regarded experts don't accept it."[62] (In order for this to spell trouble for those of us who think some propositions are absolutely indefeasible, we must take the "could" in Field's quotation to mean something like "should.") If Field is right, then it's mistaken for me to assume that *I am not in excruciating pain right now* is indefeasible, and to use this assumption to conclude that the Complex View is false.

In response, I ask the reader to consider Field's claim carefully. Is it really true that *any* proposition could rationally be called into question by expert testimony against it? Take, for example, the proposition that I exist. Ought I to diminish my credence in that proposition in the event that well-regarded experts claim otherwise? I doubt it. If I take myself to be receiving evidence from experts, must I not take myself to exist? Consider

[62] See also Quine's 1951 famous claim that "no statement is immune to revision" and Kitcher's 1983 objections to the apriorist program.

also the proposition that there are experts. Could it ever be reasonable to diminish one's credence in that proposition in light of evidence that one takes to be *from experts*? Surely any evidence that one gains from experts ought to *increase* one's credence in the proposition that there are experts, not diminish it. It's not the case, then, that expert testimony should always reduce our confidence. And I can see no other source of evidence that would fare better. (Not even testimony from God would convince me that I don't exist, for example.) I conclude, then, that some propositions are certain for us; we do occasionally enjoy absolute indefeasibility. And as far as I can tell, the proposition that *I am not in excruciating pain right now* is in that category. Therefore, it is not illegitimate for me to wield this assumption as I have in this paper, as part of an argument for the conclusion that the Complex View is false.

9. Conclusion

In this paper, I have developed and defended a novel argument for the conclusion that I am not a brain, a body, or any object with parts. Philosophers who wish to maintain the standard materialist account of human persons on which introspection is mechanistic must square their shoulders and accept the unpalatable consequence that introspection is hyperfallible, i.e.

that we can never be certain of any introspective belief. For the rest of us: introspection is not hyperfallible, and so introspection is not mechanistic. And surely at least some of our introspective beliefs are immune to defeat. Therefore the Complex View is false.

REFERENCES

Adams, Robert (1987). Flavors, Colors, and God. In *The Virtue of Faith and Other Essays in Philosophical Theology*. Oxford University Press.

Armstrong, D.M. (1963). Is Introspective Knowledge Incorrigible? *The Philosophical Review* 72: 417-32.

Armstrong, David (1968). The Headless Woman Illusion and the Defense of Materialism. *Analysis* 29: 48-49.

Baier, Kurt (1962). Smart on Sensations. *Australasian Journal of Philosophy* 40: 57-68.

Barnett, David (2005). The Problem of Material Origins. *Nous* 39: 529-540

Barnett, David (2010). You Are Simple. In G. Bealer and R. Koons (Eds.), *The Waning of Materialism*. Oxford University Press.

Barnett, David (ms.). This Wooden Table Could Have Been Made from Plastic.

Bealer, George (1992). The Incoherence of Empiricism. *Proceedings of the Aristotelian Society* Supp. Vol. 66: 99–138

Bealer, George (1996). A Priori Knowledge and the Scope of Philosophy. *Philosophical Studies* 81: 121-142

Block, Ned and Stalnaker, Robert (1999). Conceptual Analysis, Dualism, and the Explanatory Gap. *The Philosophical Review* 108: 1-46

Block, Ned (2006). Max Black's Objection to Mind-Body Identity. In Dean Zimmerman (Ed.), *Oxford Studies in Metaphysics, Vol. III.* Oxford University Press.

Burge, Tyler (1993). Mind-Body Causation and Explanatory Practice. In J. Heil and A. Mele (Eds.), *Mental Causation*. Oxford Clarendon Press.

Byrne, Alex (2005). Introspection. *Philosophical Topics* 33: 79–104.

Chalmers, David (1996). *The Conscious Mind*. Oxford University Press.

Chalmers, David (2003). The Content and Epistemology of Phenomenal Belief. In Q. Smith and A. Jokic (Eds.), *Consciousness: New Philosophical Perspectives*. Oxford University Press.

Chalmers, David (2010). *The Character of Consciousness*. Oxford University Press.

Chisholm, Roderick (1957). *Perceiving.* Princeton University Press.

Chisholm, Roderick (1978). Is There a Mind-Body Problem? *Philosophical Exchange* 2: 24-34

Chisholm, Roderick (1991). On the Simplicity of the Soul. *Philosophical Perspectives* 5: 157-181

Christensen, David (2007). Epistemology of Disagreement: The Good News. *Philosophical Review* 116: 187-217

Dennett, Daniel (1992). *Consciousness Explained*. New York: Back Bay Books.

Descartes, René (1984). *The Philosophical Writings of Descartes*, vol. II. J. Cottingham, R. Stoothoff, and D. Murdock (Eds.). Cambridge University Press.

Elga, Adam (2007). Reflection and Disagreement. *Nous* 41: 478-502.

Feldman, Richard (2006). Puzzles about Disagreement. In S. Hetherington (Ed.), *Epistemology Futures*. Oxford: Oxford University Press, 216-36

Feyerabend, Paul (1963a). Mental Events and the Brain. *Journal of Philosophy* 40:295-6

Feyerabend, Paul (1963b). Materialism and the Mind-Body Problem. *Review of Metaphysics* 17: 49-66

Fiala, Brian, Adam Arico, and Shaun Nichols (forthcoming). On the Psychological Origins of Dualism: Dual-Process Cognition and the Explanatory Gap.

Field, Hartry (2000). A Priority as an Evaluative Notion. In P. Boghossian and C. Peacocke (Eds.) *News Essays on the a Priori*. Oxford University Press.

Gennaro, R. J. (1996). *Consciousness and Self-Consciousness*. John Benjamin Publishers.

Goldman, Alan H. (2004). Epistemological Foundations: Can Experiences Justify Beliefs? *American Philosophical Quarterly* 41: 273-285

Hill, Christopher (1996). Imaginability, Conceivability, Possibility, and the Mind-Body Problem. *Philosophical Studies* 87:61-85

Horgan, Terry and Uriah Kriegel (2007). Phenomenal Epistemology: What Is Consciousness That We May Know It So Well? *Philosophical Issues* 17: 123-144.

Huemer, Michael (2007). Compassionate Phenomenal Conservatism. *Philosophy and Phenomenological Research* 74: 30-55.

Jackson, Frank (1973). Is There a Good Argument against the Incorrigibility Thesis? *Australasian Journal of Philosophy* 51: 51-62.

Kelly, Thomas (forthcoming). Peer Disagreement and Higher Order Evidence. In R. Feldman and T. Warfield (eds.), *Disagreement*. Oxford: Oxford University Press.

Kim, Jaegwon (1989). Mechanism, Purpose, and Explanatory Exclusion. In J. Tomberlin (Ed.), *Philosophical Perspectives* 3: 77-108

Kim, Jaegwon (1993). *Supervenience and Mind*. Cambridge University Press.

Kim, Jaegwon (1998). *Mind in a Physical World*. Bradford Books.

Kim, Jaegwon (2005). *Physicalism or Something Near Enough*. Princeton University Press.

Kitcher, Philip (1983). *The Nature of Mathematical Knowledge*. Oxford University Press.

Kriegel, Uriah (2007). The Same-Order Monitoring Theory of Consciousness. *Synthesis Philosophica* 44: 361-384

Kripke, Saul (1972). *Naming and Necessity*. Harvard University Press.

Lewis, David (1972). Psychophysical and Theoretical Identifications. *Australasian Journal of Philosophy* 50: 249-258

Loar, Brian (2003). Qualia, Properties, Modality. *Philosophical Issues* 13: 113-129

Lycan, William (2003). Dretske's Ways of Introspecting. In B. Gertler (Ed.), *Privileged Access*. Ashgate.

Lycan, William (2010). Giving Dualism Its Due. *Australasian Journal of Philosophy* 87: 551-563

Malcolm, Norman (1968). The Conceivability of Mechanism. *Philosophical Review* 77: 45–72.

McGinn, Colin (2003). What Constitutes the Mind-Body Problem? *Philosophical Issues* 13: 148-162

Melnyk, Andrew (1994). Being a Physicalist: How and (More Importantly) Why. *Philosophical Studies* 74: 221–41

Melnyk, Andrew (2003). *A Physicalist Manifesto: Thoroughly Modern Materialism*. Cambridge University Press.

Montero, Barbara (2001). Varieties of causal closure. In S. Walker and H.D. Heckmann (Eds.), *Physicalism and Mental Causation*. Exeter: Imprint Academic, 173-187

Nagel, Thomas (1998). Conceiving the Impossible and the Mind-Body Problem. *Philosophy* 73: 337-352

Natsoulas, T. (1996). The Case for Intrinsic Theory: I. An Introduction. *Journal of Mind and Behavior* 17: 267–286.

Nichols, Shaun and Stephen Stitch (2003). How to Read Your Own Mind: A Cognitive Theory of Self-Consciousness. In Q. Smith and A. Jokic (Eds.), *Consciousness: New Philosophical Perspectives*. Oxford University Press.

Papineau, David (1989). Why Supervenience? *Analysis* 49: 66-71.

Papineau, David (1993). The Antipathetic Fallacy. *Australasian Journal of Philosophy* 71: 169-183

Papineau, David (2002) *Thinking about Consciousness*. Oxford University Press.

Peacocke, Christopher (1979). *Holistic Explanation: Action, Space, Interpretation*. Oxford University Press.

Plantinga, Alvin (2000). Pluralism: A Defense of Religious Exclusivism. In P. Quinn and K. Meeker (Eds.), *The Philosophical Challenge of Religious Diversity*. Oxford University Press, 172-92

Polger, Thomas (2004). *Natural Minds*. MIT Press.

Pollock, John L. (1974). *Knowledge and Justification.* Princeton University Press.

Pollock, John (1986). *Contemporary Theories of Knowledge*. Rowman and Littlefield.

Pollock, John (2001). Nondoxastic Foundationalism. In M. DePaul (Ed.), *Resurrecting Old-Fashioned Foundationalism*. Rowman and Littlefield, 41-58

Quine, Willard Van Orman (1951). Two Dogmas of Empiricism. *The Philosophical Review* 60: 20-43

Rea, Michael (2002). *World without Design: The Ontological Consequences of Naturalism*. Oxford University Press.

Rorty, Richard (1965). Mind-Body Identity, Privacy and Categories. *Review of Metaphysics* 19:25-54

Rosenthal, David (2004). Varieties of Higher-Order Theory. In R.J. Gennaro (Ed.), *Higher-Order Theories of Consciousness.* John Benjamin Publishers, 19-44

Shoemaker, Sydney (1990). First-Person Access. *Philosophical Perspectives* 4:187-214

Shoemaker, Sydney (1994). Self-Knowledge and 'Inner Sense'. *Philosophy and Phenomenological Research* 54: 249-314

Shoemaker, Sydney (2001). Introspection and Phenomenal Character. *Philosophical Topics* 28:247-73

Smart, J.J.C. (1959). Sensations and Brain Processes. *The Philosophical Review* 68: 141-156

Spencer-Smith, Richard (1995). Reductionism and Emergent Properties. *Proceedings of the Aristotelian Society* 95: 113-129

Soames, Scott (2006). The Philosophical Significance of the Kripkean Necessary *Aposteriori*. *Philosophical Topics* 16:288-309

Stich, Stephen (1991). Do True Believers Exist? *Aristotelian Society Supplement* 65: 229-44.

Stich, Stephen (1996). *Deconstructing the Mind*. Oxford University Press.

Stoljar, Daniel (2009). Physicalism. In E. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy* URL = <http://plato.stanford.edu/archives/fall2009/entries/physicalism/>.

Tooley, Michael and Alvin Plantinga (2008). *Knowledge of God*. Blackwell.

Tye, Michael (1986). The Subjective Qualities of Experience. *Mind* 95: 1-17

Tye, Michael (1995). *Ten Problems of Consciousness*. MIT Press.

Tye, Michael (1999). Phenomenal Consciousness: The Explanatory Gap as Cognitive Illusion. *Mind* 108: 705-725

Tye, Michael (2000) *Consciousness, Color, and Content*. MIT Press.

Tye, Michael (2006). Another Look at Representationalism about Pain. In Murat Aydede (Ed.), *Pain: New Essays on its Nature and the Methodology of its Study*. MIT Press.

Tye, Michael (2007). Intentionalism and the Argument from No Common Content. *Philosophical Perspectives* 21: 589-613

Van Inwagen, Peter (2009). Why a Mind-Body Problem? Video interview with Robert L. Kuhn for *Closer to Truth* URL = <http://www.closertotruth.com/video-profile/Why-a-Mind-Body-Problem-Peter-van-Inwagen-/150>

Weatherson, Brian (2004). Luminous Margins. *Australasian Journal of Philosophy* 82: 373-383.

Yablo, Stephen (1992). Mental Causation. *The Philosophical Review* 101: 245-280

Zimmerman, Dean (2010). From Property Dualism to Substance Dualism. *Aristotelian Society Supplementary Volume* 84: 119-150